



COPY

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Botstein, et al.)
)
Appl. No. : Listed Below)
)
Filed : Listed Below)
)
For : SECRETED AND)
TRANSMEMBRANE)
POLYPEPTIDES AND)
NUCLEIC ACIDS ENCODING)
THE SAME)
)
Examiner : UNKNOWN)

ASSOCIATE POWER OF ATTORNEY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

The registrants of Knobbe, Martens, Olson & Bear, LLP, 2040 Main Street, Fourteenth Floor, Irvine, CA 92614, Telephone No.: (949) 760-0404, Customer No. 30,313 are hereby appointed an associate attorney to prosecute the applications listed below and take all actions respecting the same.

Please address all future communications relative to the pending patent applications listed below to the following address:

AnneMarie Kaiser
KNOBBE, MARTENS, OLSON & BEAR, LLP
2040 Main Street
Fourteenth Floor
Irvine, CA 92614
(949) 760-0404

Appl. No. : Listed Below
Filed : Listed Below

<u>Appln. No.</u>	<u>Appln. Date</u>	<u>Docket No.</u>
09/866034	5/25/2001	GNE.2930R1C1
10/033167	12/27/2001	GNE.2930R1C10
10/692072	10/22/2003	GNE.2930R1C10C1
10/213535	8/7/2002	GNE.2930R1C13
10/033244	12/27/2001	GNE.2930R1C2
10/032996	12/27/2001	GNE.2930R1C3
10/033396	12/27/2001	GNE.2930R1C4
10/033223	12/27/2001	GNE.2930R1C9

Please use Customer No. 30,313 for all communications.

Respectfully submitted,

GENENTECH, INC.

Dated: 11-02-2004

By: Elizabeth M. Barnes
Elizabeth M. Barnes, Ph.D., J.D.
Registration No. 35,059

Title: Patent Counsel

Address: 1 DNA Way
South San Francisco, CA 94080

S:\DOCS\BSG\BSG-1563.DOC
102504

GENES V

Benjamin Lewin

OXFORD UNIVERSITY PRESS
Oxford New York Tokyo
1994

~10×). Also, some tumor lines have amplified *ras* genes. A 20-fold increase in the level of a nontransforming Ras protein is sufficient to allow the transformation of some cells. The effect has not been fully quantitated, but it suggests the general conclusion that oncogenesis depends on overactivity of Ras protein, and is caused either by increasing the amount of protein or (probably more efficiently) by a variety of mutations that increase the activity of the protein.

Transfection by DNA can be used to transform only certain cell types. Although transforming oncogenes have been isolated from both rodent and

human cells, most targets for transformation by transfection with oncogenes have been rodent fibroblasts in culture. (In fact, the difference in the source of the oncogene [human] and the recipient cell [rodent] is an important factor in allowing the donor gene to be distinguished unequivocally from recipient DNA.) Limitations of the assay explain why relatively few oncogenes have been detected by transfection. This system has been most effective with *ras* genes, where there is extensive correlation between mutations that activate *c-ras* genes in transfection and the occurrence of tumors.

Insertion, translocation, or amplification may activate proto-oncogenes

A variety of genomic changes can activate proto-oncogenes, sometimes involving a change in the target gene itself, sometimes activating it without changing the protein product. In cases of insertion and translocation, there is evidence that the genomic change is the causative event; in cases of amplification there is a correlation with tumorigenesis, but no direct proof for a causative role.

Many tumor cell lines have visible regions of chromosomal amplification, as shown by homogeneously staining regions (see Figure 36.27) or double minute chromosomes (see Figure 36.28). In some cases, the amplified region contains a known oncogene or a gene related to one. In other cases, where amplification is not visible, the use of batteries of probes representing oncogenes shows that a particular oncogene is amplified. Examples of oncogenes that are amplified in various tumors include *c-myc*, *c-abl*, *c-myb*, *c-erbB*, and *c-K-ras*.

Established cell lines are prone to amplify genes (it is one of several karyotypic changes to which they are susceptible). All the same, the presence of known oncogenes in the amplified regions, and the consistent amplification of particular oncogenes in many independent tumors of the same type, again

strengthens the correlation between increased expression and tumor growth. Of course, it is possible that the gene amplification gives an advantage to growth of the established tumor; it is not necessarily an event involved in its initiation.

Some proto-oncogenes are activated by events that change their expression, but which leave their coding sequence unaltered. The best characterized is *c-myc*, whose expression is elevated by several mechanisms. One common mechanism is the insertion of a nondefective retrovirus in the vicinity of the gene.

The ability of a retrovirus to transform without expressing a *v-onc* sequence was first noted during analysis of the bursal lymphomas caused by the transformation of B lymphocytes with avian virus. Similar events occur in the induction of T cell lymphomas by murine leukemia virus. In each case, the transforming potential of the retrovirus seems to lie with its LTR rather than with a coding sequence.

In many independent tumors, the virus has integrated into the cellular genome within or close to the *c-myc* gene. The gene consists of three exons; the first represents a long nontranslated leader, and the second two code for the c-Myc

protein. Figure 59.8 summarizes the types of insertion at this locus. The retrovirus may be inserted at a variety of locations relative to the *c-myc* gene.

The simplest insertions to explain are those that occur within the first intron. The LTR provides a promoter, and transcription reads through the two coding exons. Transcription of *c-myc* under viral control differs in two ways from its usual expression: the level of expression is increased (because the LTR provides an efficient promoter); and the transcript lacks its usual nontranslated leader (which may usually limit expression).

Activation of *c-myc* in the other two classes of insertions reflects different mechanisms. The retroviral genome may be inserted within or upstream of the first intron, but in reverse orientation, so that its promoter points in the wrong direction. Probably the LTR provides an enhancer that acts on an upstream sequence that fortuitously resembles a promoter. The retroviral genome also may be inserted downstream of the *c-myc* gene, in which case transcription initiates at the usual *c-myc* promoter(s), but is increased by the enhancer in the retroviral LTR.

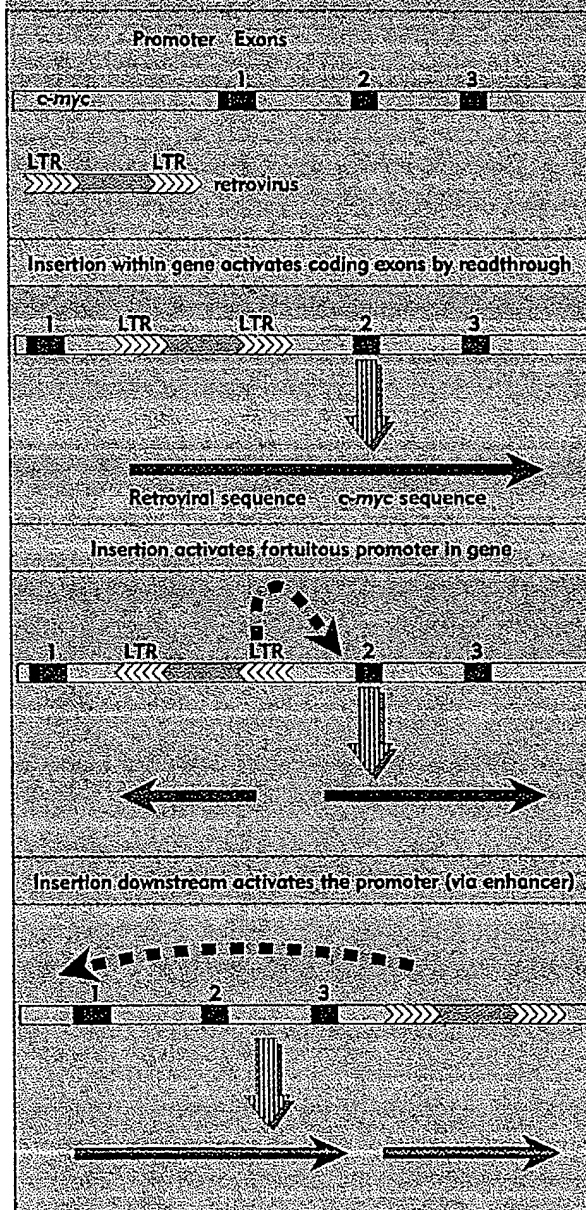
In all of these cases, the coding sequence of c-myc is unchanged, so oncogenicity is attributed to the loss of normal control, and increased expression, of the gene.

Other oncogenes that are activated in tumors by the insertion of a retroviral genome include *c-erbB*, *c-myb*, *c-mos*, *c-H-ras*, and *c-ras*. Up to 10 other cellular genes (not previously identified as oncogenes by their presence in transforming viruses) are implicated as potential oncogenes by this criterion. The best characterized among this latter class are *wnt1* and *int2*. The *wnt1* gene codes for a protein involved in early embryogenesis that is related to the *wingless* gene of *Drosophila*; *int2* codes for a growth factor related to FGF.

Translocation to a new chromosomal location is another of the mechanisms by which oncogenes are activated. Certain chromosomal translocations are consistently associated with activation of oncogenes that lie near the breakpoints. This situation was originally discovered via a connection between the loci coding immunoglobulins and the

Figure 39.8

Insertions of ALV at the *c-myc* locus occur at various positions, and activate the gene in different ways.



occurrence of certain tumors. Specific chromosomal translocations are often associated with plasmacytomas in the mouse and with Burkitt lymphomas in man. These tumors arise from undifferentiated B lymphocytes. The common feature in both species is that an oncogene on one chromosome is brought into the proximity of an Ig locus on another chromosome. Similar events occur in T lymphocytes to bring oncogenes into the proximity of a TcR locus.

The basic cause of the translocation event is a malfunction of the system responsible for recombining V and C genes or switching between IgH C genes. Instead of acting on two sites within the Ig or TcR locus, the system recombines the immune locus with an unrelated region on a different chromosome. This results in a reciprocal translocation, which is illustrated in Figure 59.9.

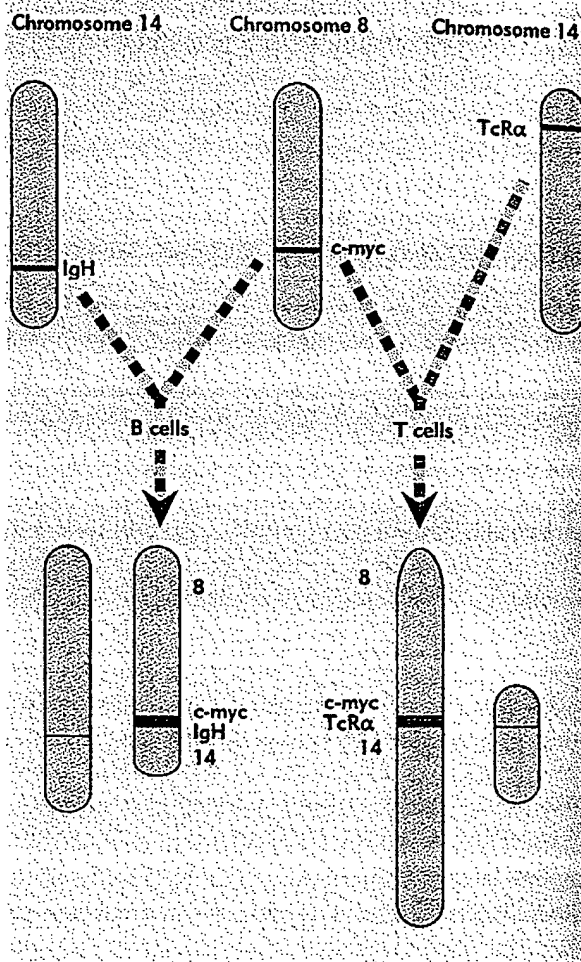
We do not know the basis for the involvement of the nonimmune partner, but in both man and mouse it is often the *c-myc* locus. In man, the translocations in B cell tumors usually involve chromosome 8, which carries *c-myc*, and chromosome 14, which carries the IgH locus; ~10% involve chromosome 8 and either chromosome 2 (κ locus) or chromosome 22 (lambda locus). The translocations in T cell tumors often involve chromosome 8, and either chromosome 14 (which has the TcR α locus at the other end from the Ig locus) or chromosome 7 (which carries TcR β locus). Analogous translocations occur in the mouse.

Translocations at the IgH locus in B cells fall into two classes. One type is similar to those observed at other Ig loci and at TcR loci, involving the consensus sequences used for V-D-J somatic recombination of active Ig genes. In the other type, the translocation occurs at a switching site, so these cases may be associated with function of the system that switches expression from one C_H gene to another.

When *c-myc* is translocated to the Ig locus, its level of expression is usually increased. The increase varies considerably among individual tumors, generally being in the range from 2 to 10. Why does translocation activate the *c-myc* gene? The translocation event does not involve fixed sites,

Figure 39.9

A chromosomal translocation is a reciprocal event that exchanges parts of two chromosomes. Translocations that activate the human *c-myc* proto-oncogene involve Ig loci in B cells and TcR loci in T cells.



but occurs at a variety of locations within a general region on each recombining chromosome. The event has two consequences: *c-myc* is brought into a new region, one in which an Ig or TcR gene was actively expressed; and the structure of the *c-myc* gene may itself be changed (but usually not involving the coding regions). It seems likely that several different mechanisms can activate the *c-myc* gene

in its new location (just as retroviral insertions activate *c-myc* in a variety of ways).

The correlation between the tumorigenic phenotype and the activation of *c-myc* by either insertion or translocation suggests that continued high expression of the c-Myc protein is oncogenic. Expression of *c-myc* must be switched off to enable immature lymphocytes to differentiate into mature B and T cells; failure to turn off *c-myc* maintains the cells in the undifferentiated (dividing) state. The oncogenic potential of *c-myc* has been demonstrated directly by the creation of transgenic mice carrying a normal *c-myc* gene linked to an enhancer. Transgenic mice carrying a *c-myc* gene linked to a B lymphocyte-specific enhancer (the IgH enhancer) develop lymphomas. The tumors represent both immature and mature B lymphocytes, suggesting that over-expression of *c-myc* is tumorigenic throughout the B cell lineage. Transgenic mice carrying a *c-myc* gene under the control of the LTR from a mouse mammary tumor virus, however, develop a variety of cancers, including mammary carcinomas. This suggests that increased or continued expression of *c-myc* transforms the type of cell in which it occurs into a corresponding tumor. Specificity of the tumor type may therefore depend on the mechanism used to activate *c-myc*; it is not an intrinsic property of the gene.

c-myc exhibits three means of oncogene activation: retroviral insertion, chromosomal translocation, and gene amplification. The common thread among them is increased expression of the oncogene rather than a qualitative change in its coding function, although in at least some cases the transcript has lost the usual (and possibly regulatory) nontranslated leader. *c-myc* provides the paradigm for oncogenes that may be effectively activated by increased (or possibly altered) expression.

Every translocation generates reciprocal products; sometimes a known oncogene is activated in one of the products, but in other cases it is not evident which of the reciprocal products has responsibility for oncogenicity. Also, it is not axiomatic that the gene(s) at the breakpoint have

responsibility; for example, the translocation could provide an enhancer that activates another gene nearby.

A variety of translocations found in B and T cells have identified new oncogenes. In some cases, the translocation generates a hybrid gene, in which an active transcription unit is broken by the translocation. This has the result that the exons of one gene may be connected to another. In such cases, there are two potential causes of oncogenicity: the proto-oncogene part of the protein may be activated in some way that is independent of the other part, for example, because it is over-expressed under its new management (a situation directly comparable with the example of *c-myc*); or the other partner in the hybrid gene may have some positive effect that generates a gain-of-function in the part of the protein coded by the proto-oncogene.

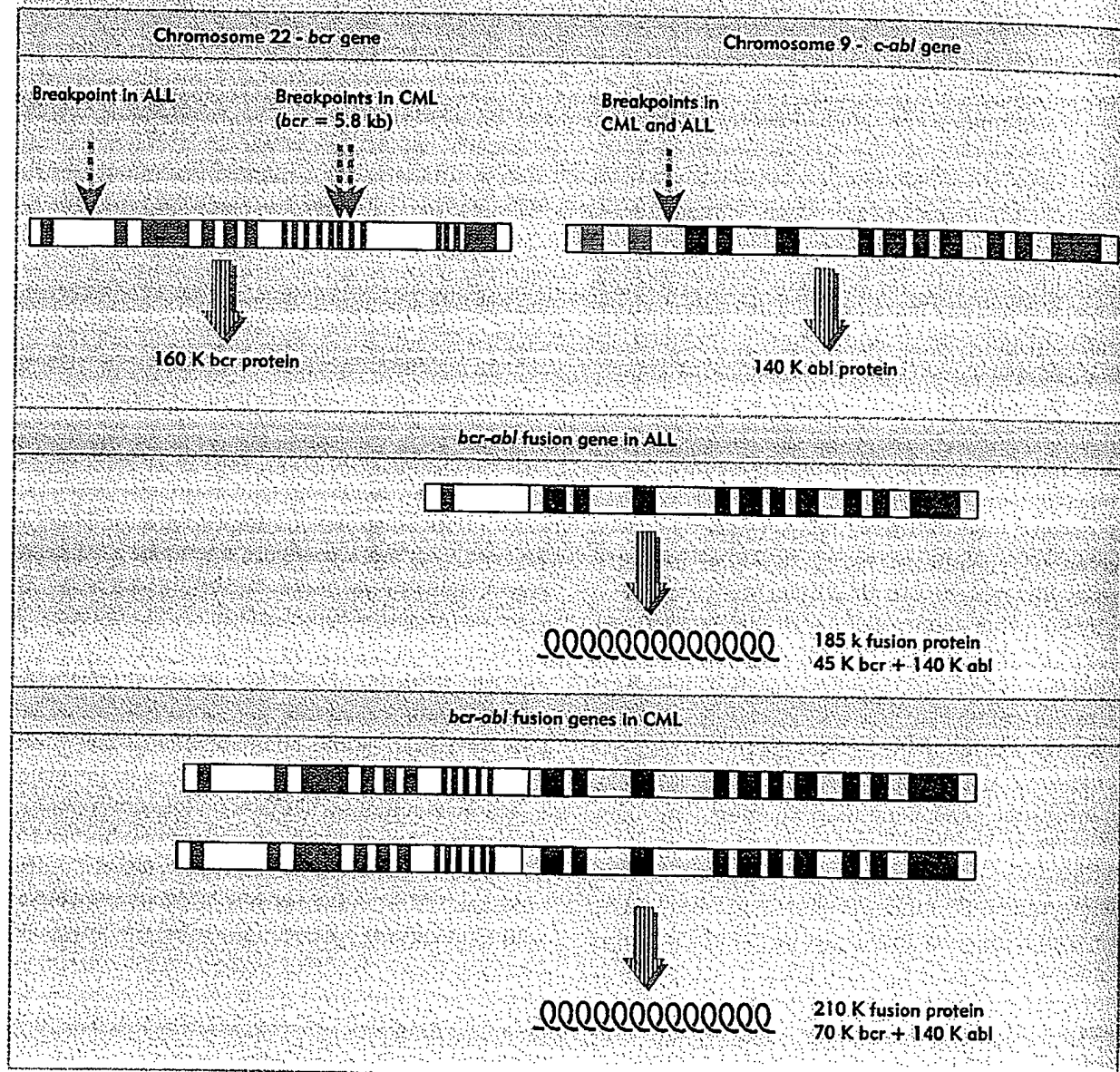
One of the best characterized cases in which a translocation creates a hybrid oncogene is provided by the Philadelphia (*PH*¹) chromosome present in patients with chronic myelogenous leukemia (CML). This reciprocal translocation is too small to be visible in the karyotype, but links a 5,000 kb region from the end of chromosome 9 carrying *c-abl* to the *bcr* region of chromosome 22. The *bcr* (break-point cluster region) was originally named to describe a region of ~5.8 kb within which breakpoints occur on chromosome 22. Different cases of CML have breakpoints at different locations within this region.

The consequences of this translocation are summarized in Figure 39.10. The *bcr* region lies within a large (>90 kb) gene, which is now known as the *bcr* gene. The breakpoints in CML usually occur within one of two introns in the middle of the gene. The same gene is also involved in translocations that generate another disease, ALL (acute lymphoblastic leukemia); in this case, the breakpoint in the *bcr* gene occurs in the first intron.

The *c-abl* gene is expressed by alternative splicing that uses either of the first two exons. The breakpoints in both CML and ALL occur in the intron that precedes the first common exon. Although the exact breakpoints on both chromosomes 9 and 22 vary in individual cases, the

Figure 39.10

Translocations between chromosome 22 and chromosome 9 generate Philadelphia chromosomes that synthesize *bcr-abl* fusion transcripts that are responsible for two types of leukemia.



common outcome is the production of a transcript coding for a Bcr-Abl fusion protein, in which N-terminal sequences derived from *bcr* are linked to *c-abl* sequences. In ALL, the 185,000 dalton

fusion protein has ~45,000 daltons of the Bcr protein linked to c-Abl; in CML the fusion protein of 210,000 daltons has ~70,000 daltons of the Bcr protein. In each case, the fusion protein contains ~140,000

daltons of the usual ~145,000 c-Abl protein, that is, it has lost just a few N-terminal amino acids of the c-Abl sequence.

Why is the fusion protein oncogenic? It relies on an interaction between the N-terminal region provided by *bcr* with the c-Abl protein. The *bcr* gene has a variety of sequence motifs related to proteins involved in signalling pathways, but the pertinent one is a serine/threonine kinase activity that is coded within the first exon. This autophosphorylates residues in this part of the protein, and the phosphorylation confers the ability to interact with a region of the c-Abl protein called the SH2 domain (we discuss the nature of SH2 domains later). This

raises the possibility that the Bcr part of the protein interacts with the c-Abl sequences, perhaps changing their conformation and activating a latent oncogenic potential.

Changes at the N-terminus are involved in the activation of oncogenic activity of *v-abl*, a transforming version of the gene carried in a retrovirus. The *c-abl* gene codes for a tyrosine kinase activity; this activity is essential for transforming potential in oncogenic variants. Deletion (or replacement) of the N-terminal region activates the kinase activity and transforming capacity. So the N-terminus provides a domain that usually regulates kinase activity; its loss may cause inappropriate activation.

Loss of tumor suppressors causes tumor formation

The common theme in the role of oncogenes in tumorigenesis is that increased or altered activity of the gene product is oncogenic. Whether the oncogene is introduced by a virus or results from a mutation in the genome, it is dominant over its allelic proto-oncogene(s). A mutation that activates a single allele is tumorigenic. Tumorigenesis then results from gain of a function.

Certain tumors are caused by a different mechanism: loss of both alleles at a locus is tumorigenic. Propensity to form such tumors may be inherited through the germline; it also occurs as the result of somatic change in the individual. Tumorigenesis then results from loss of function. Such cases identify tumor suppressors: genes whose products are needed for normal cell function, and whose loss of function causes tumors. The two best characterized genes of this class code for the proteins RB and p53.

Retinoblastoma is a human childhood disease, involving a tumor on the retina. It occurs both as a heritable trait and sporadically (by somatic mutation). It is often associated with deletions of band q14 of human chromosome 13. The *RB* gene has been localized to this region by molecular cloning.

Figure 39.11 illustrates the situation.

Retinoblastoma arises when both copies of the *RB* gene are inactivated. In the inherited form of the disease, one parental chromosome carries an alteration in this region, usually a deletion. A somatic event in retinal cells that causes loss of the other copy of the *RB* gene causes a tumor. In the sporadic form of the disease, the parental chromosomes are normal, and both *RB* alleles are lost by (individual) somatic events.

Almost half the cases of retinoblastoma show deletions at the *RB* locus. In other cases, transcripts of the locus are either absent or altered in length. The protein product is absent from retinoblastoma cells. The cause of the tumor is therefore loss of protein function, usually resulting from mutations that prevent gene expression (as opposed to point mutations that affect function of the protein product). Loss of *RB* is involved also in other forms of cancer, including osteosarcomas and small cell lung cancers.

What is the molecular function of RB? It interacts with a variety of other proteins, including several tumor antigens: SV40 T antigen, adenovirus E1A, human papilloma virus E7. One possibility is that part of the oncogenicity of these proteins is due to

AMPLIFICATION OF CELLULAR ONCOGENES IN CANCER CELLS

K. ALITALO

FROM THE DEPARTMENT OF VIROLOGY, UNIVERSITY OF HELSINKI, HELSINKI, FINLAND

ABSTRACT

Regulatory or structural alterations of cellular oncogenes have been implicated in the causation of various cancers. Oncogene alteration by point mutations can result in a protein product with strongly enhanced oncogenic potential. Aberrant expression of cellular oncogenes may be due to tumour-specific chromosomal translocations that dysregulate the normal functions of a proto-oncogene. Amplification of cellular oncogenes can also augment their expression by increasing the amount of DNA template available for the production of mRNA. It appears that amplification of certain oncogenes is a common correlate of the progression of some tumours and also occurs as a rare sporadic event affecting various oncogenes in different types of cancer. Amplified copies of oncogenes may or may not be associated with chromosomal abnormalities signifying DNA amplification: double minute chromosomes and homogeneously staining chromosomal regions. Amplified oncogenes, whether sporadic or tumour type-specific, are expressed at elevated levels, in some cases in cells where their diploid forms are normally silent. Increased dosage of an amplified oncogene may contribute to the multistep progression of at least some cancers.

KEY WORDS: CELLULAR ONCOGENES, GENE AMPLIFICATION, MULTISTEP CARCINOGENESIS, CLONAL SELECTION, KARYOTYPIC ABNORMALITIES, DOUBLE MINUTE CHROMOSOMES, HOMOGENEOUSLY STAINING CHROMOSOMAL REGIONS

DNA SEQUENCE AMPLIFICATION AND CYTOGENETIC ABNORMALITIES IN TUMOURS

Since its discovery in drug-resistant eukaryotic cells, somatic amplification of specific genes has been implicated in an increasing variety of adaptive responses of cells to environmental stresses (70, 79). Cytogenetic abnormalities, double minute chromosomes (dmin:s) associated with DNA amplification had already been discovered in tumour cells before the discovery of dmin:s and homogeneously staining chromosomal regions (HSR:s) in cells selected for drug-resistance (12, 24, 49, 50, 56). In metaphase spreads, dmin:s appear as small, spherical, usually paired chromosome — like structures that lack a centromere (Fig. 2). HSR:s stain with intermediate intensity throughout their length rather than with the normal pattern of alternating dark and light bands in both trypsin-Giemsa (Fig. 3A) and quinacrine dihydrochloride-stained preparations. Both kinds of abnormalities are occasionally found in metaphases of freshly isolated cancer cells but not of normal cells (8).

Dmin:s and HSR:s are apparently rare in tumour cells in vivo, although exact data are

difficult to obtain since the abnormalities are easily missed in routine cytogenetic analysis (8, 42). Dmin:s and HSR:s have been described in most types of in vitro-cultured malignant tumour cells, with a notable frequency in neuroblastoma cell lines (11). Initial growth in cell culture apparently selects for tumour cells that contain either dmin:s or HSR:s. Moreover, in culture dmin:s are frequently lost, concomitant with the appearance of clonal populations of cells that have developed an HSR, suggesting that the two cytogenetic abnormalities are alternative forms of gene amplification and that HSR:s may confer a selective advantage on cells over dmin:s (11, 70). It has been assumed that HSR:s can break down to firm dmin:s and that dmin:s can integrate into chromosomes to generate HSR:s (11, 23). Amplified genes may also occupy abnormally banding regions, ABR:s (51, 59). Experimental work on drug-resistant cells has shown that in the absence of a selection pressure (drug), dmin:s and the amplified genes that they contain are lost, whereas amplified DNA in the form of HSR:s is retained in the cells (71). This is explained by the fact that dmin:s are segregated unevenly in mitosis and frequently get lost from the nucleus due to

ONC
Retro
lexan

RSV
Y73V
GR-F
Ab-M
FuSV

ST-an
GA-F
UR2V
AEV

SM-F

MH-2
3911-
Mo-M
SSV

Ha-M
K-MS

FBJ-A
OK-1
AMV

SKV 7
REV
AEV
E26V

ONCC
Tumour

Neuro
Neuro
Small
Neuro

their
some
divid
HSR:
grow
follo
form
grow
and
ly in
totox
HSR:
any f
that
possi
(By d

TABLE 1
Currently known oncogenes.

ONCOGENES FOUND IN RETROVIRUSES				
Retrovirus (example)	Oncogene	Gene product		
		Cellular location	Function of protein	Class
RSV	<i>src</i>	Plasma membrane	Tyrosine-specific protein kinases (<i>fgr</i> contains sequences homologous to actin)	Class 1a (Cytoplasmic tyrosine protein kinases)
Y73V	<i>yes</i>	Plasma membrane		
GR-FeSV	<i>fgr</i>	Plasma membrane		
Ab-MuLV	<i>abl</i>	Plasma membrane		
FuSV	<i>fps/fes</i>	Cytoplasm (plasma membrane?)		
ST-and GA-FeSV	<i>fes/fps</i>	Cytoplasm (cytoskeleton?)		
UR2V	<i>ras</i>			
AEV	<i>erb-B</i>	Plasma membrane and cytoplasmic membranes	EGF receptor's cyto- plasmic domain	Class 1b (Class 1a-related proteins)
SM-FeSV	<i>fms</i>	Plasma membrane and cytoplasmic membranes	Cytoplasmic domain of a growth factor receptor?	
MH-2V	<i>mil/raf</i>	Cytoplasm	?	
3911-MSV	<i>raftmil</i>	Cytoplasm	?	
Mo-MSV	<i>mos</i>	Cytoplasm	?	
SSV	<i>sis</i>	Secreted	PDGF-like growth factor	Class 2 (Growth factors)
Ha-MSV	<i>Ha-ras</i>	Plasma membrane	GTP-binding proteins	Class 3 (Cytoplasmic GTP:ases)
Ki-MSV	<i>Ki-ras</i>	Plasma membrane		
FBJ-MuSV	<i>fos</i>	Nucleus	?	Class 4 (Nuclear phospho- proteins)
OK-10V	<i>myc</i>	Nucleus	Nuclear matrix protein	
AMV	<i>myb</i>	Nucleus	?	
SKV 770	<i>ski</i>	Nucleus?	?	Unclassified
REV	<i>rel</i>	?	?	
AEV	<i>erb-A</i>	?	?	
E26V	<i>ets</i>	?	?	
ONCOGENES FOUND IN TUMOUR CELLS BUT NOT IN RETROVIRUSES				
Tumour cell				
Neuroblastoma	<i>N-ras</i>	Plasma membrane	GTP-binding	Class 3
Neuroblastoma	<i>N-myc</i>	?	?	Class 4
Small-cell lung cancer	<i>L-myc</i>	?	?	Class 4
Neuro-/Glioblastomas	<i>neu</i>	Plasma membrane	Growth factor receptor	Class 1b

their lack of centromeres, (49). HSR chromosomes carry centromeres and are therefore divided equally at mitosis. If dmin:s and HSR:s contain amplified genes that encode growth-stimulating protein products, it would follow that the more stable chromosomal form, the HSR, confers a greater selective growth advantage for cells. Although dmin:s and HSR:s have been described predominantly in tumour cells selected for resistance to cytotoxic drugs, it is also clear that dmin:s and HSR:s may be present in cancer cells before any form of therapy (8). It was in this setting that we and others first chose to explore the possible amplification of cellular oncogenes. (By definition, cellular oncogenes are normally

innocent genetic loci which can be activated to transforming genes in various ways).¹

DMIN:S AND HSR:S CONTAIN AMPLIFIED ONCOGENES

Table 2 summarizes the somatic amplifications of cellular oncogenes so far reported in

¹ It is not the purpose of this review to deal with all forms of DNA damage that have been found to activate cellular oncogenes. For the purpose of integrating the review into a coherent picture, however, the reader is given a list of known cellular oncogenes in Table 1 and the schematic Figure 1 illustrating the various ways in which the oncogenic potential of different proto-oncogenes can be activated.

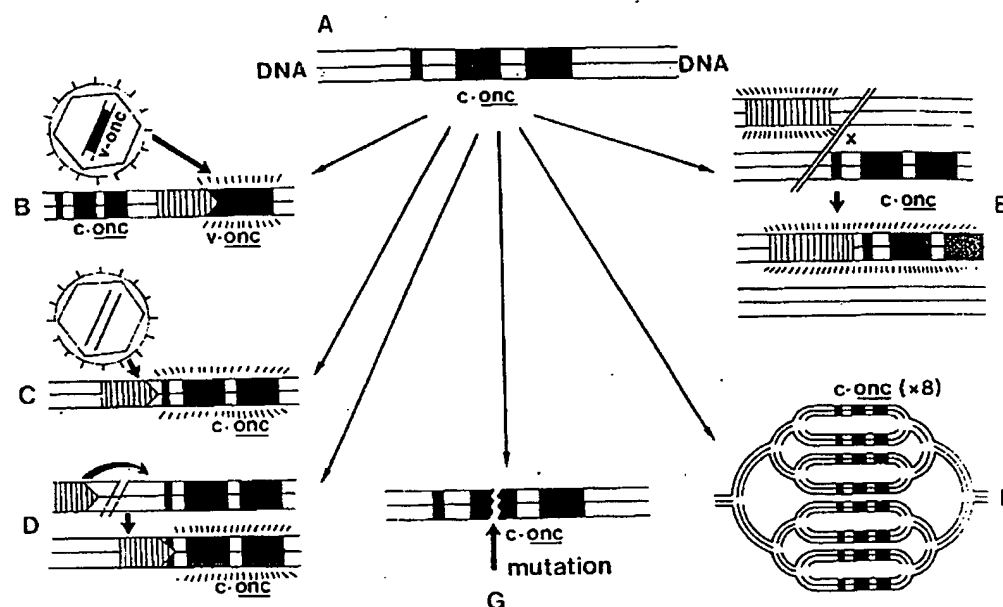


Fig. 1. Activation of cellular oncogenes. The haploid complement of a proto-oncogene is schematically depicted in A, composed of three exons (black boxes) in a segment of DNA. The different activated forms are schematically outlined in B–G. The abbreviation *c-onc* stands for cellular oncogene, *v-onc* viral oncogene, DNA sequences with associated strong promoter/enhancer functions are striated, and an actively transcribed gene is marked with radiations. B. Acute transforming retroviruses have the capacity to transduce cellular oncogenes (*c-onc*) into their genome, modify them and reinsert their activated oncogenes (*v-onc*) into the genome of host animal cells as a part of their proviral forms. The activity of the *v-onc* gene is greatly enhanced due to the associated promoter of the proviral long terminal repeat. Both increased dosage of the oncogene and structural mutations within its sequence may contribute to tumorigenesis. C. Slow transforming retroviruses without oncogenes replicate and reinsert their proviral copies into the host cell DNA during a latency period from infection to tumorigenesis. Tumor initiation through hyperplastic growth may begin, when the provirus integrates sufficiently close to a proto-oncogene to activate it through promoter or enhancer functions. It should be noted, however, that mutations have also been found in the resulting tumors. D. In some mouse plasmacytomas, a retrovirus-like DNA element (directing the synthesis of the so-called intracisternal A-type particles, IAPs) has been found in association with a transcriptionally activated oncogene *c-mos*. The IAP insertion also disrupts the 5' part of *c-mos* [64]. E. In humans, as well as in animals, chromosome translocations may place proto-oncogenes into transcriptionally active regions of chromatin, where they may be activated. The details of this mechanism have not been worked out, but it is believed to occur for *c-myc* and *c-abl* genes in Burkitt lymphomas and Philadelphia-chromosome positive leukemias, respectively [35, 40]. F. Increased amounts of oncogene-specific RNA and protein can also result from an excess of DNA template for transcription acquired through oncogene amplification. The present review concentrates primarily on this mechanism. G. Mutationally activated oncogenes have been found in nearly one fifth of human malignant tumours. Oncogene loci activated by somatic structural mutations are revealed by transfection experiments, where they are introduced into genetic background of nontumorigenic cultured immortalized cells. Several such transforming loci have been cloned and many of them belong to the *c-ras* oncogene family. It should be pointed out that both structural mutations and either increased expression or activation of a complementing oncogene may be required to achieve a fully tumorigenic phenotype [44].

tumour cells. Although the sampling of tumours is at present small, the finding of known cellular oncogenes among amplified DNA represented by *dmin:s* and *HSR:s* of cancer cells is provocative. Amplification has been found to affect at least five out of twenty known cellular oncogenes and the degree of gene amplification varies from five to many hundred-fold over the single haploid copies found in normal cells (see also ref. 18). The first amplification reported involved the *c-myc*

oncogene (see Table 1) in a promyelocytic leukaemia cell line HL-60 [20, 25]. The degree of *c-myc* amplification is between 8–32 fold both in the HL-60 cell line and in primary leukaemic cells from the patient [20, 25]. Original clonal lines of HL-60 were later found to contain some *dmin:s* in culture but their number was insufficient to establish any clear correlation with amplified *c-myc*. Such a correlation, however, was discovered for *c-myc* amplification in a neuroendocrine cell line from

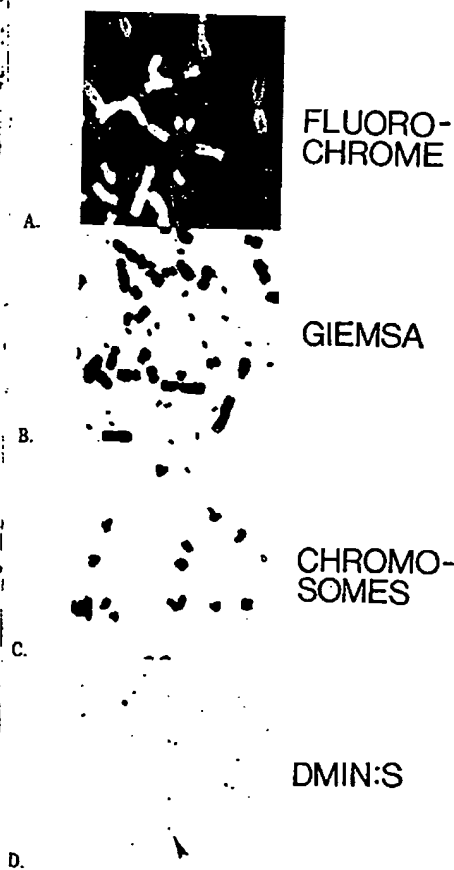


Fig. 2. Double minute chromosomes (arrowheads) in COLO 320DM colon carcinoma cells. A. The dmin:s are resolved as paired dots among normal chromosomes in this fluorescent, benzimidazole-stained preparation B—D. Purification of dmin:s by differential centrifugations. B. The starting material, C. Chromosome fraction, D. Purified dmin:s (Donna George and the author, unpublished data and ref. 52).

a colon carcinoma, COLO 320 (5). In these cells, the approximately 30-fold amplified *c-myc* copies were mapped either to HSR:s of a marker chromosome (5, Fig 3B) or to dmin:s (52), depending on the particular subline studied. Since dmin:s were already present in the primary tumour cells from this colon carcinoma (63), it is likely that *c-myc* had also been amplified during growth of the tumour in vivo. Similarly, amplified copies of the *c-Ki-ras* oncogene were mapped to dmin:s and HSR:s of a mouse adrenocortical tumor Y1 (74). An extensive search for changes in other oncogenes and tumour cells has since revealed amplifications that do not show up as dmin:s or HSR:s.

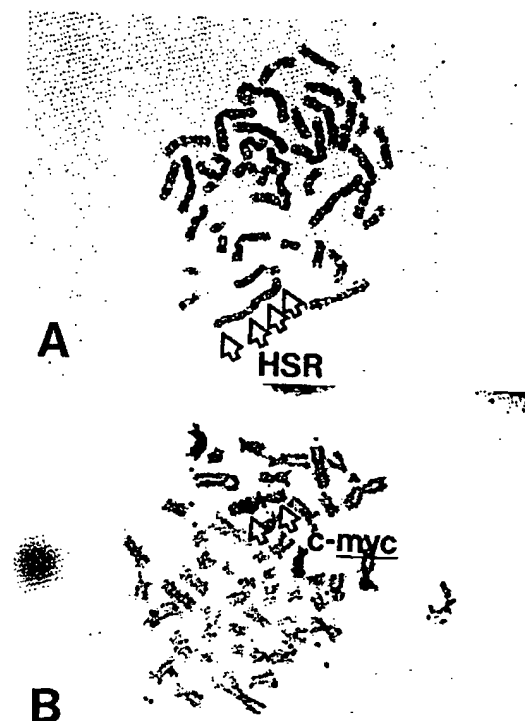


Fig. 3. A. The homogeneously staining regions (HSR) in the G-banded HSR-marker chromosome comprise a major portion of both its long and short arms. The HSR-marker chromosome has evolved from an X-chromosome (52 and unpublished data of C. C. Lin and the author). B. The about 30-fold amplified copies of the *c-myc* oncogene in COLO 320 cells were found to be located to dmin:s and HSR:s. The latter is shown here by in situ-hybridization (5, 52).

Thus, for example, the *c-myc* oncogene is amplified in a characteristic marker chromosome of a colon carcinoma without evidence of HSR:s (ref. 6, Fig. 4) and in other tumours, the amplified *c-abl* and *c-myc* oncogene loci map to abnormally banding regions (ABR:s) in translocated or resident chromosomal segments, respectively (59, 76).

TRANSLOCATIONS AND REARRANGEMENTS MAY ACCOMPANY ONCOGENE AMPLIFICATION

The evolution and progression of the karyotype of tumour cells is complicated (see ref. 68). Concomitant with amplification, DNA sequences acquire an increased mobility in the genome with extrachromosomal intermediates

TABLE 2

*Sporadic and tumour-specific amplification of cellular oncogenes.**

Tumour cells	Oncogene	Fold amplified	Chromosomal location of amplified gene	Expression elevated	Remarks	References
Sporadic:						
HL60 (acute promyelocytic leukaemia, M3)	<i>c-myc</i>	20x	8q(ABR)	Yes	Amplification present in primary leukaemic cells	20, 25, 59
COLO320 (colon carcinoma)	<i>c-myc</i>	30x	dmin, HSR	Yes	Part of the amplified <i>c-myc</i> sequences rearranged	4, 5, 52
Y1 (adrenocortical tumour)	<i>c-Ki-ras</i>	50x	dmin, HSR	Yes	Levels of p 21 ^{c-Ki-ras} protein elevated	74
COLO201/205 (colon carcinoma)	<i>c-myb</i>	10x	mar1	Yes	Patient treated with 5-fluorouracil prior to culturing of the tumour cells	4, 6, 88
K562 (chronic myelogenous leukaemia, CML)	<i>c-abl</i>	10x	mar(ABR)	Yes	C ₁ coamplified in the marker that may be derived from chromosome 22, <i>c-abl</i> protein-associated tyrosine kinase activated	21, 22, 41, 54, 76
A431 (epidermoid carcinoma)	<i>c-erbB</i>	15–20x	n.d.	Yes	Amplification linked to chromosome 7 translocation and sequence rearrangements. Amount of protein product, the EGF receptor, elevated	82 (see 36)
ML1–3 (acute myeloid leukaemia, M2)	<i>c-myb</i>	5–10x	n.d.	Yes	Abnormalities of chromosome 6q22–24, where <i>c-myb</i> is normally located	34, 61, 91
SK BR-3 (breast carcinoma)	<i>c-myc</i>	10x	n.d.	Yes		43
SEWA (polyoma virus-induced mouse tumour)	<i>c-myc</i>	30x	n.d.	Yes	Cells have dmin:s depending on culture conditions; <i>c-myc</i> amplification correlates with growth as a tumour	Manfred Schwab, personal communication
Lu-65 (lung giant cell carcinoma)	<i>c-myc</i>	8x	n.d.	n.d.	At least some copies of <i>c-Ki-ras</i> mutated	80
	<i>c-Ki-ras</i>	10x	n.d.	n.d.		
Primary leukemic cells from an acute myeloid leukemia (M2) patient	<i>c-myc</i>	33x	n.d.	n.d.		Unpublished data of the author and A. de la Chapelle
Tumour-specific:						
small-cell lung cancer	<i>c-myc</i> , <i>L-myc</i> , <i>N-myc</i>	up to 80x	n.d.	Yes	Most amplifications in the variant phenotype of SCLC	53, 69
Neuroblastomas	<i>N-myc</i>	up to 250x	dmin, HSR	Yes	<i>N-myc</i> also amplified in primary tumours of advanced grade	14, 48, 72, 73, 75
Glioblastomas	<i>c-erbB</i>	—	—	—	Rearrangements also found	Josef Schlessinger, personal communication

n.d. = not determined, mar = marker chromosome, M2, M3 refer to the French-American-British classification of acute myeloid leukemias.

* At least one case of oncogene amplification in normal germ-line cells has been found (18).

visualized as dmin:s, transpositions and translocations to other chromosomal segments, etc. (see 70 for references). There may not be preferred chromosomal sites for the apparent reintegration of dmin:s as HSR:s (75). In at

least one case, however, an oncogene may have been caught amplifying in situ in its resident chromosomal site (59). The finding of moderately amplified oncogenes also in chromosomal sites lacking HSR:s suggests that



Fig. 4. Localization of amplified *c-myc* in COLO 201/205 cells by in situ hybridization. Shown is a characteristic, large marker chromosome (mar1) with G-banding (A) and associated *c-myc* autoradiographic grains (B). Note the absence of HSRs. Mar1 has probably evolved from chromosome number 6, the resident site of the *c-myc* oncogene in normal cells [34, 88, 91]. (Robert Winqvist and the author, unpublished data).

[onco]gene amplification may be more common than the structural alterations shown by chromosome banding and microscopy [6, 88].

In at least three cases reported amplification has been accompanied by a DNA rearrangement of the oncogene [5, 20, 82]. In the colon carcinoma COLO 320 both damaged and normal versions of the *c-myc* gene are amplified [5]. Although individual cell clones have not yet been examined, our unpublished experiments suggest that the same dmin-containing cells harbor and express both normal and rearranged forms of *c-myc*. The normal version of the amplified gene, however, predominates in COLO 320 cells containing HSRs; the rearranged version is present only in what appears to be a single copy (Fig. 5). In the chronic myeloid leukaemia (erythroleukemia) cell line K562 an amplified DNA segment consists of portions of both the *c-abl* oncogene and the immunoglobulin C_λ locus [76]. In both cases abnormal transcripts are produced from the rearranged amplified oncogenes (Fig. 6 and ref. 22). In K562 cells, the abnormal *c-abl* oncogene product has also been activated as a tyrosine protein kinase [41]. It is not known, however, whether structural alterations of the genes preceded amplification or whether they were acquired during the process of gene amplification. It seems likely that a chromosomal translocation of *c-abl* to the C_λ locus preceded DNA amplification in the K562 cells, since all amplified copies were also rearranged [21], with the change reminiscent of the Philadelphia translocation [t(9, 22)]

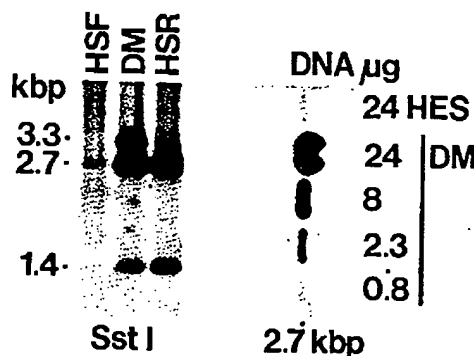


Fig. 5. Amplification and rearrangement of *c-myc* in COLO 320 cells. 10 μ g of cellular DNA was digested with Sst I, electrophoresed, blotted and probed with a *v-myc* Pst I fragment (ref. 2, left panel). Fragments of 2.7 kbp and 1.4 kbp are seen in both normal and amplified *c-myc* DNA. The 3.3 kbp fragment is derived from a DNA segment of unknown origin translocated to the 5' region of *c-myc* with a concomitant deletion of its first exon (unpublished data of Manfred Schwab and the author). HSF, human skin fibroblasts; DM, COLO 320 DM cells; HSR, COLO 320 HSR cells. Different amounts of DNA from COLO 320 DM cells as indicated were mixed with calf thymus DNA to give 24 μ g of total DNA, cleaved with Sst I, electrophoresed, blotted and probed with a fragment of 3' human *c-myc* sequences. The intensities of the 2.7 kbp *c-myc* fragment in different samples were compared to assess its copy number, estimated to be about 30 [5].

found in most chronic myeloid leukaemia tumours [35, 66–68]. Although they have not been sequenced, other reported cases of amplified oncogenes are apparently normal on basis of mapping with restriction endonucleases (see Table 2). Therefore we cannot at present view mutation as a necessary companion of oncogene amplification.

THE MECHANISMS OF GENE AMPLIFICATION

The mechanisms of gene amplification and the structure of the amplified DNA have been worked out mainly in experimental settings involving selection for drug-resistance in cell culture [70]. Although the mechanisms are still incompletely known and may vary in different cases, some general features have emerged.

A spontaneous degree of illegitimate DNA replication seems to exist in normal cells so that various segments of DNA are replicated more than once during a single cell cycle [37]. In unselective conditions this DNA is probably lost e.g., through formation of micronucleae

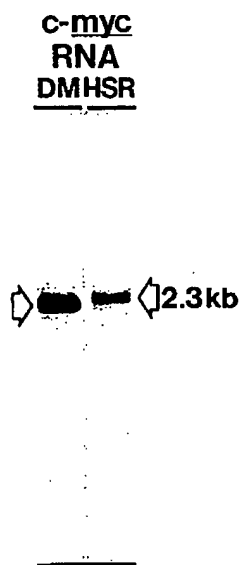


Fig. 6. Comparison of the electrophoretic mobilities of *c-myc* mRNA:s from COLO 320 DM and HSR cells. The size of the normal *c-myc* mRNA is 2.3 kb. The rearranged *c-myc* locus in DM cells (see Fig. 5) seems to be predominantly expressed giving rise to a shortened RNA.

because the newly synthesised extra copies of DNA are not covalently linked to chromosomal DNA of mitotic cells (65, 71). If, however, there is a selection pressure to retain an increased gene dosage, progressive multiplication of gene copy number results. The incidence of cells bearing amplified genes under conditions of cytotoxic selection can vary by two orders of magnitude and is greatly increased by the presence of mitogenic substances (hormones or tumour promoters) during selection (10, 84, 85) or certain carcinogenic or cytotoxic agents before selection (15, 55, 79, 80, 81, 85). An interesting hypothesis suggested by Varshavsky (84, 85) supposes that the origins of DNA replication "fire" (initiate replication) illegitimately several times during a single cell cycle and that this kind of "replicon misfiring" may be increased by substances such as tumour promoters and mitogenic hormones (10, 84, 85). Mariani and Schimke (55) point out that most of the cytotoxic agents that increase the incidence of gene amplification are inhibitors of DNA synthesis. Aberrant replication is known to take place after transient inhibition of DNA synthesis and this response can lead to gene amplification (46, 47, 55, 90). Mitogenic hormones probably increase disproportionate DNA replication, but they

also enhance the colony forming efficiency of drug-resistant cells in selective conditions (10).

According to the studies of Axel and his collaborators (65), the multiple cycles of unscheduled DNA replication at a single locus during a single cell cycle result in a structure schematically outlined in Fig. 1F. The hydrogen-bonded amplified copies of DNA depicted in Fig. 1F must resolve into a tandem linear array before the next mitosis. This may well occur by homologous recombination between any one of several repeated sequences within the amplified domain (45, 65). Part of the recombinations would lead to extrachromosomal circles possessing an origin for replication (16, 62); these could be the precursors of *dmin:s*. The unequal recombinations mean that the resolved linear structure consists of tandemly repeated but heterogeneous units. According to Axel's model a gradient of amplification is formed so that centrally located sequences are amplified more than sequences distal to the origin of replication (65). This also has, in fact, been found to explain the large, complex DNA domain amplified in neuroblastoma cells *in vivo* (38, see also below).

The chromosomal site of integration of transfected genes significantly affects the frequency and cytogenetic result of their experimentally induced amplification (83). The amplification frequency in some transformants has been found to be 100-fold that of the others (83). This suggests that there also are preferred chromosomal positions for amplification of host cellular genes and that chromosomal rearrangements may facilitate gene amplification by positioning chromosomal sequences in a favorable array. In respect of the structural properties of the sequences involved in gene amplification, recombinatorially active regions have been implicated in experimental cases. DNA rearrangements involving restriction fragment length polymorphisms and variation in gene copy number have been detected in the human genome between clusters of short repetitive interspersed DNA sequences called Alu family DNA-sequences (17). Such inter-Alu sequences have also been detected in an extrachromosomal DNA form, including covalently closed circles (17, 78). The copy number of inter-Alu sequences apparently varies in an age- and tissue-specific manner (17, 78), but any comprehensive analysis of the phenomenon in human tumours is not yet available. It is also not yet clear whether these kinds of repetitive sequences are involved in generating amplified oncogene sequences in *dmin:s* or HSR:s in tumours.

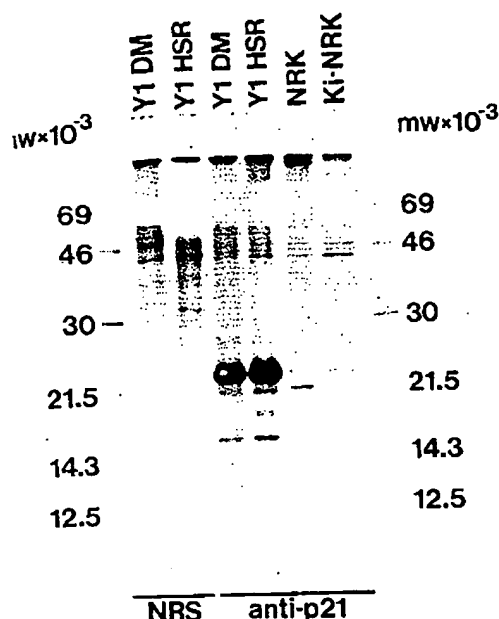


Fig. 7. Elevated levels of the p21^{Ki-ras} protein in Y1 cells (74). The Y1 DM and HSR cells which harbor a 50-fold amplified c-Ki-ras oncogene (74) and control cells were labeled with [³⁵S]-methionine and the p21^{Ki-ras} protein was immunoprecipitated, as detailed (74), with normal rat serum (NRS) or rat monoclonal anti-p21 serum. The proteins were electrophoresed in a 15 % polyacrylamide gel in the presence of SDS. In addition to a major p21 band, a labeled band at about 16 kd was present in the immunoprecipitates. The amount of radioactivity in p21 was about 50 fold that in normal rat kidney cells. The Kristen sarcoma virus-transformed rat kidney cells (obtained from the American Tissue Culture Collection) also yielded unexpectedly low amounts of the v-Ki-ras protein.

CARCINOGEN-INDUCED GENE AMPLIFICATION AND CLONAL SELECTION OF CANCER CELLS

Although cell sorting experiments have shown a basal spontaneous rate of gene amplification in eukaryotic cells (37), this can be increased severalfold by metabolic inhibitors or cytotoxic agents (15, 37, 70, 81, 85). In many respects the latter response is reminiscent of the so-called SOS-response elicited in bacteria by noxious stimuli (see 28). In a teleological context, the rapid induction of gene amplification that apparently occurs frequently through extrachromosomal intermediates may provide cells with genetic material for subsequent selective pressures operating in harmful conditions (60). In cancer cells, the mechanism may enhance the emergence of

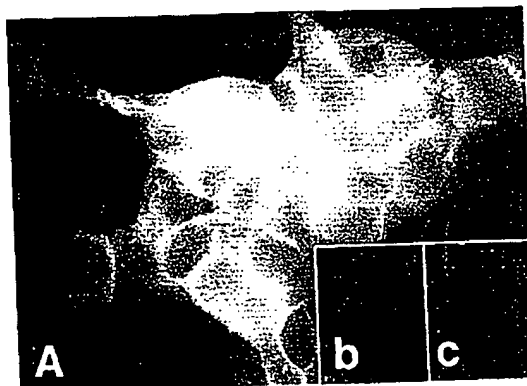


Fig. 8. A. Indirect immunofluorescence for p21^{Ki-ras} in Y1 DM cells. Similar fluorescence of the plasma membrane was obtained for the Y1 HSR cells. Inset (b) shows control staining with normal rat serum and inset (c) staining of normal rat kidney cells with the monoclonal antibody against p21.

clonal populations of cells with increasingly malignant properties (58, 60). Such genetic instability of cancer cells is clearly enhanced, leading to the rapid evolution of increasingly malignant tumour cell populations (19, 58). A serious question of practical importance is whether drug resistance in treated patients also selects cells that have an enhanced ability to amplify (onco)genes important for growth and progression of the tumour (84, 85). It is also possible that some of the carcinogenic insults caused by mutagens are only expressed as a result of subsequent amplification events induced by tumour promoters (84, 85) or facilitated by hormones in, say replicating epithelial cells (10). The persistence of dmin:s in some tumours suggests that there is a selection pressure for their retention (8, 9, 11, 23). Amplified DNA in dmin:s must contain an origin for DNA replication (62) and must be selected for in daughter cell populations, where it is unevenly segregated (71). In the absence of such a selection pressure dmin:s are lost (71). In at least one study the length of an HSR has been found to increase during a selection of malignant cells for enhanced tumorigenicity (30).

The amplified c-erbB gene in A431 cells codes for epidermal growth factor receptor (27). The abundant amounts of receptor protein on A431 cell surface may, however, provide the cells with an abnormal growth response (31). A naive supposition is that the amplified sequences in dmin:s and possibly in HSR:s of tumours contain growth-promoting genes (see 36 for references). This seems to fit well with

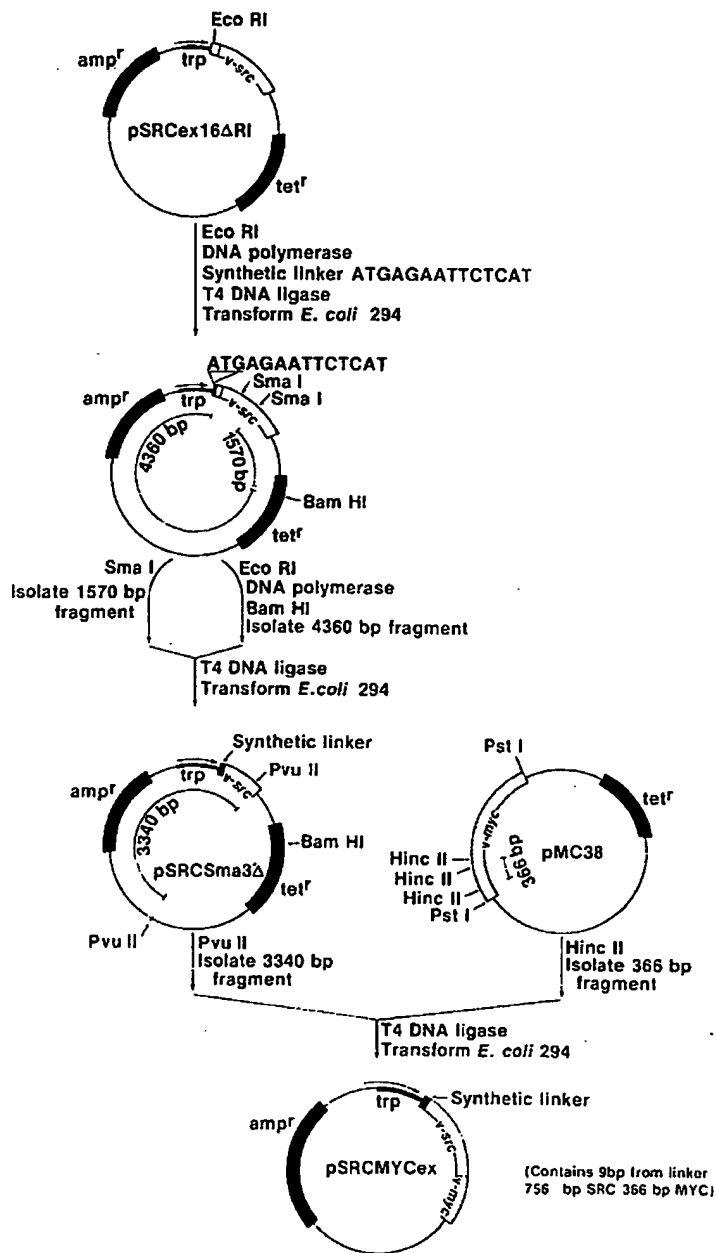


Fig. 9. Construction of a v-myc expression vector. A synthetic linker [ATGAGAATTCATCAT] containing a translational initiation codon was inserted downstream from the *trp* promoter in the pSRC ex16 RI expression vector described previously (see ref. 3). Approximately one-half of the *v-src* sequences coding for the aminoterminal portion of pp60^{src} protein were then deleted and the remaining portion ligated in translational codon frame with the synthetic ATG. A Hinc II fragment of *v-myc* from plasmid clone MC 38 (nucleotides 320–685 in the *v-myc* sequence in ref. 2) was ligated downstream from remaining *v-src* sequences in continuity with its reading frame. The resulting product contained 3 amino acids from the synthetic linker, 252 amino acids encoded by the 756 base pair fragment from Sma I to Pvu II restriction sites in *v-src* DNA, 122 amino acids from the *v-myc* and 6 amino acids [corresponding to nucleotides 2968–2085] from the pBR322 vector (3).

recent findings on amplified oncogenes, though in many cases the search for an amplified oncogene is still continuing. Even positive findings do not mandate a role for amplified cellular oncogenes, however, because the domain of amplified DNA is inevitably much larger than a single genetic locus (e.g. 38).

ENHANCED EXPRESSION OF AMPLIFIED ONCOGENES

In all cases where they have been studied, the amplified oncogenes have been found abundantly expressed at the RNA level, roughly in proportion to the amount of DNA amplification (see Table 1). Described cases of elevated RNA expression include examples of abnormal (5, 22) and ectopic (6) transcription. In at least four cases this enhancement is not limited to synthesis of RNA (31, 33, 41, 74, 82). The Y1 cells that have amplified *c-Ki-ras* contain exceptionally large amounts of its protein product situated on the plasma membrane (ref. 74, Fig. 7 and 8). High amounts of the *c-myc* encoded protein are also found in COLO 320 cells that have amplified the gene (33). The *myc* oncogenes have recently been shown to encode nuclear proteins (ref. 1, 3, 26, 29, 32, 33, Fig. 9–11). Both the expression of the *c-myc* mRNA (39) and the subcellular localization of *myc* proteins are linked to the cell cycle (ref. 89, Fig. 12). It may be that elevated expression of specific *c-myc* functions is necessary for cell cycle progression and the growth transformation aspect of the phenotype of cancer cells that may contribute to tumour progression (7, 36). Elevated expression of *c-myc* has been shown to replace in part platelet-derived growth factor in induction of competence for DNA replication (7). Generally, enhanced expression of an oncogene could be a necessary prerequisite for acquisition of a growth advantage by cells having extra copies of the gene. This effect

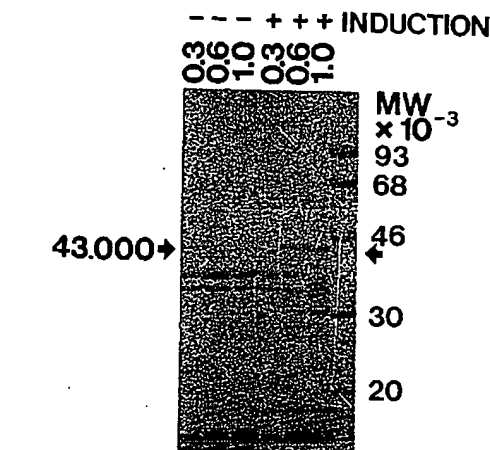


Fig. 10. *E. coli* 294 was transfected with the hybrid *v-src*, *v-myc* plasmid outlined in Fig. 9 and ampicillin-resistant bacterial colonies were checked for the production of a 43,000 m.w. bacterial *v-myc* protein after induction by growth to different optical densities in minimal essential medium (M9, induction +) or complete medium (LB, induction-) (3).

could also be the principal contribution of amplification to tumourigenesis.

TUMOUR CELL AND STAGE SPECIFICITY OF ONCOGENE ACTIVATION AND AMPLIFICATION

Tumour cell specificity of oncogene amplification has been found in three malignancies. The *c-myc*, *L-myc* or *N-myc* oncogene is amplified in most cases of the variant form of small-cell lung cancer cells (53, 69), *c-erbB* is amplified in several glioblastomas (Josef Schlessinger, personal communication) and the putative *N-myc* oncogene is amplified in about half of grade III and IV neuroblastomas (14, 72, 73, 75). In addition to HSR:s, small-cell lung cancers and neuroblastomas frequently show a



Fig. 11. Indirect immunofluorescence for the *v-myc* protein and phase contrast microscopy of myelocytomatosis virus-transformed quail cells (3). A. Quail cells transformed with the MC-29 virus (Q8 cells). Anti-*myc* protein staining. B. Phase contrast microscopy of field in A. C. Q8 cells stained with anti-*myc* protein antiserum that has been blocked with the immunogen. D. Phase contrast microscopy of field in B. E. Q8 cell stained with preimmune rabbit serum. F. Phase contrast microscopy of field in E.

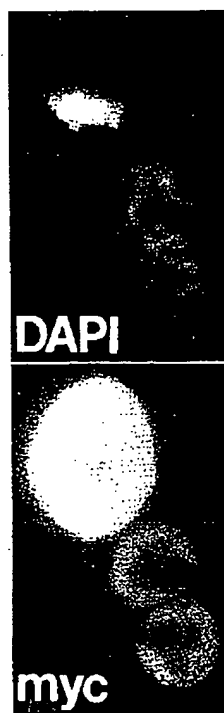


Fig. 12. Fluorescent staining for DNA and *myc* protein in myelocytomatosis virus-transformed quail cells. In interphase cells, the *myc* protein is confined to the nucleus. In the mitotic cell, *myc* fluorescence is distributed throughout the cell unlike fluorescence for chromatin, which is compacted to chromosomes in the metaphase plate. In fact, there is less *myc* fluorescence associated with chromatin than with the rest of the cell. DAPI, diamidophenylindole DNA stain. The anti-*myc* protein rabbit antiserum was used in a 1/200 dilution [ref. 89].

deletion of a portion of the short arm of chromosome 1 [13] and chromosome 3 [86, 87], respectively, in karyological examination. Two kinds of changes have also been described in different neuroblastoma oncogenes. The first is a mutation in the *N-ras* gene, an activated oncogene that was discovered because of its relation to other *ras* genes and transforming activity in transfection experiments [77]. The second is amplification of a distant homologue of the *c-myc* gene called *N-myc* [72, 73, 75]. Although the transforming potential of the *N-myc* gene has not yet been established, its consistent presence in a core segment of amplified neuroblastoma DNA [38, 57, 72, 73, 75] and its elevated expression in most retinoblastomas [48] suggests its oncogenic nature.

Taya et al. [80] have recently described a human lung giant cell carcinoma grown in nude mice, where both *c-Ki-ras* and *c-myc* on-

cogenes were amplified about 10-fold. Besides, sequencing studies indicated that at least some of the amplified *c-Ki-ras* copies were also mutationally activated in the 12th codon. These results fit to the multistage theory of cancer development and progression [see 58]. Apparently co-operating lesions in cellular oncogenes accumulate during tumour growth and selection and increase the malignant potential of the tumour cells [44].

When does oncogene amplification come in to play during tumorigenesis? Gene amplification may not be any initiating event in carcinogenesis. Amplification and enhanced expression of *c-myc* and *N-myc* may occur during the progression of human carcinoma of the lung and neuroblastoma cells to a more malignant phenotype [14, 53, 73]. There may be, however, no mandatory sequence of oncogene amplifications for the genesis of any particular tumor. Amplification of an oncogene could play its part in malignant progression of already initiated cells whenever it happened to occur.

ACKNOWLEDGEMENTS

I thank my colleagues Manfred Schwab, Gerard Evan, J. Michael Bishop, Robert Winqvist, Kalle Saksela, Jorma Keski-Oja, C. C. Lin, Arthur Levinson, Wendy Colby and Donna George for collaboration, Ron Ellis for monoclonal antibodies against p21 proteins and Stephen Hann and Robert Eisenman for communicating their results before publication. The studies in the author's laboratory were supported by the Finnish Cancer Research Fund and by the Academy of Finland.

REFERENCES

1. Abrams HD, Rohrschneider LR, Eisenman RN: Nuclear location of the putative transforming protein of avian myelocytomatosis virus. *Cell* 29: 427-439, 1982
2. Alitalo K, Bishop JM, Smith DH, Chen EY, Colby WW, Levinson AD: Nucleotide sequence of the *v-myc* oncogene of avian retrovirus MC29. *Proc Natl Acad Sci USA* 80: 100-104, 1983
3. Alitalo K, Ramsay G, Bishop JM, Ohlsson-Pfeifer S, Colby WW, Levinson AD: Identification of nuclear proteins encoded by viral and cellular *myc* oncogenes. *Nature* 306: 274-277, 1983
4. Alitalo K, Saksela K, Winqvist R, Schwab M, Bishop JM: Amplification and aberrant expression of cellular oncogenes in human colon cancer cells. In: *Genes and cancer*. Ed. J. M. Bishop and J. Rowley. Alan Liss Co, New York, in press
5. Alitalo K, Schwab M, Lin CC, Varmus HE, Bishop JM: Homogeneously staining chromosomal regions con-

- tain amplified copies of an abundantly expressed cellular oncogene (*c-myc*) in malignant neuroendocrine cells from a human colon carcinoma. *Proc Natl Acad Sci USA* 80: 1707-1711, 1983
6. Alitalo K, Winqvist R, Lin CC, de la Chapelle A, Schwab M, Bishop JM: Aberrant expression of an amplified *c-myc* oncogene in two cell lines from a colon carcinoma. *Proc Natl Acad Sci USA* 81: 4534-4538, 1984
 7. Armelin HA, Armelin MCS, Kelly K, Stewart T, Leder P, Cochran BH, Stiles CD: Functional role for *c-myc* in mitogenic response to platelet-derived growth factor. *Nature* 310: 655-660, 1984
 8. Barker PE: Double Minutes in human tumor cells. *Cancer Genet Cytogenet* 5: 81-94, 1982
 9. Barker PE, Drwinga HL, Mittelman WN, Maddox AM: Double minutes replicate once during S phase of the cell cycle. *Exp Cell Res* 130: 353-360, 1980
 10. Barsom J, Varshavsky A: Mitogenic hormones and tumor promoters greatly increase the incidence of colony-forming cells bearing amplified dihydrofolate reductase genes. *Proc Natl Acad Sci USA* 80: 5330-5334, 1983
 11. Biedler JL, Meyers MB, Spengler BA: Homogeneously staining regions and double minute chromosomes, prevalent cytogenetic abnormalities of human neuroblastoma cells. *Adv Cell Neurobiol* 4: 268-301, 1983
 12. Biedler JL, Spengler BA: Metaphase chromosome anomaly: association with drug resistance and cell-specific products. *Science* 191: 185-187, 1976
 13. Brodeur GM, Green AA, Hayes FA, Williams KJ, Williams DL, Tsatis AA: Cytogenetic features of human neuroblastomas and cell lines. *Cancer Res* 41: 4678-4686, 1981
 14. Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM: Amplification of *N-myc* in Untreated Human Neuroblastomas Correlates with Advanced Disease Stage. *Science* 224: 1121-1124, 1984
 15. Brown PC, Tlsty TD, Schimke RT: Enhancement of methotrexate resistance and DHFR gene amplification by treatment of 3T6 cells with hydroxyurea. *Mol Cell Biol* 3: 1097-1107, 1983
 16. Bullock P, Botchan M: Molecular events in the excision of SV40 DNA from the chromosomes of cultured mammalian cells. In: *Gene amplification*. Ed. R. T. Schimke. Cold Spring Harbor Lab 215-230, 1982
 17. Calabretta B, Roberson DL, Barrera-Saldana HA, Lambrou TP, Saunders GF: Genome instability in a region of human DNA enriched in Alu repeat sequences. *Nature* 296: 219-225, 1982
 18. Chattopadhyay SK, Chang EH, Lander MR, Ellis RW, Scalnick EM, Lowy DR: Amplification and rearrangement of *onc* genes in mammalian species. *Nature* 296: 361-363, 1982
 19. Cifone MA, Fidler IJ: Increasing metastatic potential is associated with increasing genetic instability of clones isolated from murine neoplasms. *Proc Natl Acad Sci USA* 78: 6949-6952, 1981
 20. Collins S, Groudine M: Amplification of endogenous *myc*-related DNA sequences in a human myeloid leukemia cell line. *Nature* 298: 679-681, 1982
 21. Collins SJ, Groudine MT: Rearrangement and amplification of *c-abl* sequences in the human chronic myelogenous leukemia cell line K-562. *Proc Natl Acad Sci USA* 80: 4813-4817, 1983
 22. Collins SJ, Kubonishi I, Miyoshi I, Groudine MT: Altered transcription of the *c-abl* oncogene in K-562 and other chronic myelogenous leukemia cells. *Science* 225: 72-74, 1984
 23. Cowell JK: Double minutes and homogeneously staining regions: Gene amplification in mammalian cells. *Ann Rev Genet* 16: 21-52, 1982
 24. Cox D, Yunchen C, Spriggs AI: Numerous minute chromatin bodies in malignant tumours of childhood. *Lancet* I: 55-58, 1965
 25. Dalla-Favera RD, Wong-Staal F, Gallo RG: *Onc* gene amplification in promyelocytic leukemia cell line HL-60 and primary leukaemic cells of the same patient. *Nature* 299: 61-63, 1982
 26. Donner P, Greiser-Wilke I, Moelling K: Nuclear Localization and DNA binding of the transforming gene product of avian myelocytomatosis virus. *Nature* 296: 262-266, 1982
 27. Downward J, Yarden Y, Mayes E, Scrace G, Totty N, Stockwell P, Ulrich A, Schlessinger J, Waterfield MD: Close similarity of epidermal growth factor receptor and *v-erbB* oncogene protein sequences. *Nature* 307: 521-527, 1984
 28. Echols H: SOS-functions, cancer and inducible evolution. *Cell* 25: 1-2, 1981
 29. Eisenman RN, Tachibana CY, Abrams HD, Hann SR: *v-myc* and *c-myc*-Encoded proteins are associated with the nuclear matrix. Submitted
 30. Gilbert F, Balaban G, Brangman D, Herrmann N, Lister A: Homogeneously staining regions and tumorigenicity. *Int J Cancer* 31: 765-768, 1983
 31. Gill GN, Lazar CS: Increased phosphotyrosine content and inhibition of proliferation in EGF-treated A431 cells. *Nature* 293: 305-307, 1981
 32. Hann SR, Abrams HD, Rohrschneider LR, Eisenman RN: Proteins encoded by *v-myc* and *c-myc* oncogenes: Identification and localization in acute leukemia virus transformants and bursal lymphoma cell lines. *Cell* 34: 789-798, 1983
 33. Hann SR, Eisenman RN: Two proteins encoded by the human *c-myc* oncogene: Differential expression in neoplastic cells. *Mol Cell Biol* 4: 2486-2497, 1984
 34. Harper ME, Franchini G, Love J, Simon MI, Gallo RC, Wong-Staal F: Chromosomal sublocalization of human *c-myc* and *c-fes* cellular onc genes. *Nature* 304: 169-171, 1983
 35. Heisterkamp N, Stephenson JR, Groffen J, Hansen PF, de Klein A, Bartman CR, Grosveld G: Localization of the *c-abl* oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. *Nature* 306: 239-242, 1983
 36. Heldin C-H, Westermark B: Growth factors: Mechanism of action and relation to oncogenes. *Cell* 37: 9-20, 1984
 37. Johnston RN, Beverley SM, Schimke RT: Rapid spontaneous dihydrofolate reductase gene amplification shown by fluorescence-activated cell sorting. *Proc Natl Acad Sci USA* 80: 3711-3715, 1983
 38. Kanda N, Schreck R, Alt F, Bruns G, Baltimore D, Latt S: Isolation of amplified DNA sequences from IMR-32 human neuroblastoma cells: Facilitation by fluorescence-activated flow sorting of metaphase chromosomes. *Proc Natl Acad Sci USA* 80: 4069-4073, 1983
 39. Kelly K, Cochran BH, Stiles CD, Leder P: Cell-specific regulation of the *c-myc* gene by lymphocyte mitogens and platelet-derived growth factor. *Cell* 35: 603-610, 1983
 40. Klein G: Specific chromosome translocations and the genesis of B-cell derived tumors in mice and men. *Cell* 32: 311-315, 1983
 41. Konopka JB, Watanabe SM, Witte ON: An alteration of the human *c-abl* protein in K562 leukemia cells unmasks associated tyrosine kinase activity. *Cell* 37: 1035-1042, 1984
 42. Kovacs G: Homogeneously staining regions on marker

- chromosomes in malignancy. *Int J Cancer* 23: 299–301, 1979
43. Kozbor D, Croce CM: Amplification of the *c-myc* oncogene in one of five human breast carcinoma cell lines. *Cancer Res* 44: 438–441, 1984
 44. Land H, Parada LF, Weinberg RA: Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Science* 222: 771–778, 1983
 45. Latt SA: Sister chromatid exchange formation. *Ann Rev Genet* 15: 11–15, 1981
 46. Laughlin TJ, Taylor JH: Initiation of DNA replication in chromosomes of Chinese hamster ovary cells. *Chromosoma* 75: 19–35, 1979
 47. Lavi S: Carcinogen-mediated amplification of viral DNA sequences in simian virus 40-transformed Chinese hamster embryo cells. *Proc Natl Acad Sci USA* 78: 6144–6148, 1981
 48. Lee W-H, Murphree AL, Benedict WF: Expression and amplification of the *N-myc* gene in primary retinoblastoma. *Nature* 309: 458–460, 1984
 49. Levan A, Manolov G, Clifford P: Chromosomes of a human neuroblastoma. A new case with accessory minute chromosomes. *J Natl Cancer Inst* 41: 1377–1387, 1968
 50. Levan A, Levan G, Mitelman F: Chromosomes and cancer. *Hereditas* 86: 15–29, 1977
 51. Lewis JA, Biedler JL, Melera PW: Gene amplification accompanies low level increases in the activity of dihydrofolate reductase in antifolate-resistant Chinese hamster lung cells containing abnormally banding chromosomes. *J Cell Biol* 94: 418–424, 1982
 52. Lin CC, Alitalo K, Schwab M, George D, Varmus HE, Bishop JM: Evolution of karyotypic abnormalities and *c-myc* oncogene amplification in a human colonic carcinoma. Submitted for publication
 53. Little CD, Nau MM, Carney DN, Gazdar AF, Minna JD: Amplification and expression of the *c-myc* oncogene in human lung cancer cell lines. *Nature* 306: 194–196, 1983
 54. Luzzio CB, Luzzio BB: Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* 45: 321–334, 1975
 55. Mariani BD, Schimke RT: Gene amplification in a single cell cycle in Chinese hamster ovary cells. *J Biol Chem* 259: 1901–1910, 1984
 56. Mark J: Double minutes—a chromosomal aberration in Rous sarcomas in mice. *Hereditas* 57: 1–22, 1967
 57. Montgomery KT, Biedler JL, Spengler BA, Melera PW: Specific DNA sequence amplification in human neuroblastoma cells. *Proc Natl Acad Sci USA* 80: 5724–5728, 1983
 58. Nowell RC: The clonal evolution of tumor cell populations. *Science* 194: 23–28, 1976
 59. Nowell R, Finan J, Favera RD, Gallo RC, Ar-Rushdi A, Romanczuk G, Selden JR, Emanuel BS, Rovera G, Croce CM: Association of amplified oncogene *c-myc* with an abnormally banded chromosome 8 in a human leukemia cell line. *Nature* 306: 494–497, 1983
 60. Pall ML: Gene amplification model of carcinogenesis. *Proc Natl Acad Sci USA* 78: 2465–2468, 1981
 61. Pelicci P-G, Lanfrancone L, Brathwaite MD, Wolman SR, Dalla-Favera R: Amplification of the *c-myc* oncogene in a case of human acute myelogenous leukemia. *Science* 224: 1117–1121, 1984
 62. Pellegrini S, Dailey L, Basilico C: Amplification and excision of integrated polyoma DNA sequences require a functional origin of replication. *Cell* 36: 943–949, 1984
 63. Quinn LA, Moore GE, Morgan RT, Woods LK: Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogeneously staining regions. *Cancer Res* 39: 4914, 1979
 64. Rechavi G, Givol D, Canaani E: Activation of a cellular oncogene by DNA rearrangement: possible involvement of an IS-like element. *Nature* 300: 607–611, 1982
 65. Roberts JM, Buck LB, Axel R: A structure for amplified DNA. *Cell* 33: 53–63, 1983
 66. Rowley JD: Human oncogene locations and chromosome aberrations. *Nature* 301: 290, 1983
 67. Rowley JD, Testa JR: Chromosome abnormalities in malignant hematologic diseases. *Adv Cancer Res* 36: 103–147, 1982
 68. Rowley JD, Ulmann JE eds: *Chromosomes and Cancer: from molecules to man*. Academic Press, New York 1983
 69. Sahsela K, Bergh J, Lehto V-P, Nilsson K, Alitalo K: Amplification of the *c-myc* oncogene is characteristic of a subpopulation of human small cell lung cancer. *Cancer Res*, in press
 70. Schimke RT: Gene amplification. Cold Spring Harbor Labor 1982
 71. Schimke RT, Brown PC, Kaufman RJ, McGrogan M, Slate DL: Chromosomal and extrachromosomal localization of amplified dihydrofolate reductase genes in cultured mammalian cells. *Cold Spring Harbor Symp Quant Biol* 55: 785–797, 1981
 72. Schwab M, Alitalo K, Klempnauer K-H, Varmus HE, Bishop JM, Gelbert F, Brodeur G, Goldstein M, Trent J: Amplified DNA with limited homology to *myc* cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* 305: 245–248, 1983
 73. Schwab M, Alitalo K, Varmus HE, Bishop JM: Amplification of cellular oncogenes in tumor cells. In: *The cancer cell*. Cold Spring Harbor Press 2: 215–220, 1984
 74. Schwab M, Alitalo K, Varmus HE, Bishop JM, George D: A cellular oncogene [*c-Ki-ras*] is amplified, overexpressed, and located within karyotypic abnormalities in mouse adrenocortical tumour cells. *Nature* 303: 497–501, 1983
 75. Schwab M, Varmus HE, Bishop JM, Grezeschik K-H, Naylor SL, Sakaguchi AY, Brodeur G, Trent J: Chromosome localization in normal human cells and neuroblastomas of a gene related to *c-myc*. *Nature* 308: 288–291, 1984
 76. Selden JR, Emanuel BS, Wang E, Cannizzaro L, Palumbo A, Erikson J, Nowell PC, Rovera G, Croce CM: Amplified *C_λ* and *c-abl* genes are on the same marker chromosome in K562 leukemia cells. *Proc Natl Acad Sci USA* 80: 7289–7292, 1983
 77. Shimizu K, Goldfarb M, Perucho M, Wigler M: Isolation and preliminary characterization of the transforming gene of human neuroblastoma cell line. *Proc Natl Acad Sci USA* 80: 383–387, 1983
 78. Shmookler Reis RJ, Lumpkin CK, McGill JR, Riabowol TK, Goldstein S: Extrachromosomal circular copies of an 'inter-Alu' unstable sequence in human DNA are amplified during *in vitro* and *in vivo* ageing. *Nature* 201: 394–398, 1983
 79. Stark GR, Wahl GM: Gene amplification. *Ann Rev Biochem* 53: 447–491, 1984
 80. Taya Y, Hesogai K, Hirohashi S, Shimozato Y, Tsuchiya R, Tsuchida N, Fushimi M, Sekiya T, Nishimura S: A novel combination of *K-ras* and *myc* amplification accompanied by point mutational activation of *K-ras* in human lung cancer. *EMBO J* 3: 2943–2946, 1984
 81. Tlsty TD, Brown PC, Schimke TR: UV radiation facilitates methotrexate resistance and amplification of the dihydrofolate reductase gene in cultured 3T6 mouse cells. *Molec Cell Biol* 4: 1050–1056, 1984

82. Ulrich A, Coussens L, Haylick JS, Dull TJ, Gray A, Tam AW, Lee J, Yarden Y, Libermann TA, Schlessinger J, Downward J, Mayes ELV, Whittle N, Waterfield MD, Seeburg PH: Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* 309: 418-425, 1984
83. Wahl GM, de Saint Vincent BR, DeRose ML: Effect of chromosomal position on amplification of transfected genes in animal cells. *Nature* 307: 516-520, 1984
84. Varshavsky A: On the possibility of metabolic control of replicon 'misfiring': Relationship to emergence of malignant phenotypes in mammalian cell lineages. *Proc Natl Acad Sci USA* 78: 3673-3677, 1981
85. Varshavsky A: Phorbol ester dramatically increases incidence of methotrexate-resistant mouse cells: Possible mechanisms and relevance to tumor promotion. *Cell* 25: 561-572, 1981
86. Whang-Peng J, Bunn PA, Kao-Shan CS, Lee EC, Carney DN, Gazdar A, Minna JD: A Nonrandom chromosomal abnormality, del 3p (14-23), in human small cell lung cancer (SCLC). *Cancer Genet Cytogenet* 6: 119-134, 1982
87. Whang-Peng J, Kao-Shan CS, Lee EC, Bunn PA, Carney DN, Gazdar AF, Portlock C, Minna JD: Deletion 3p (14-23), Double minute chromosomes, and homogeneously staining regions in human small-cell lung cancer. *sCSH Laboratory* 1982
88. Winqvist R, Knuutila S, Leprince D, Stehelin D, Alitalo K: Mapping of amplified *c-myc* oncogene, sister chromatid exchanges and karyotypic analysis of the COLO 205 colon carcinoma cell line. *Cancer Genet Cytogenet.* in press
89. Winqvist R, Saksela K, Alitalo K: *myc* proteins are not associated with chromatin in mitotic cells. *EMBO J* 3: 2947-2950
90. Woodcock DM, Cooper IA: Evidence for double replication of chromosomal DNA segments as a general consequence of DNA replication inhibition. *Cancer Res* 41: 2483-2490, 1981
91. Zabel BU, Naylor SL, Grezeschik K-H, Sakaguchi AY: Regional assignment of human protooncogene *c-myc* to 6q21-qter. *Somatic Cell Molec Genet* 10: 105-108, 1984

Received for publication: September 19, 1984

Address: K. Alitalo
Department of Virology
University of Helsinki
SF-00290 Helsinki
Finland

mus HE,
M, Trent
myc cel-
blastoma
ure 305:

Amplifi-
In: The
15-220,

I, George
ed, over-
bnormal-
s. *Nature*

chik K-H,
Chromo-
nd neuro-
ture 308:

L, Pahlm-
roce CM:
the same
ells. *Proc*

M: Isola-
the trans-
cell line.
93

Riabowol
r copies of
DNA are
og. *Nature*

Anr Rev

Y, Ts...hiya
imura S: A
mplification
on of K-ras
1946, 1984
iation faci-
ification of
ltured 376
156, 1984

Elevated Epidermal Growth Factor Receptor Gene Copy Number and Expression in a Squamous Carcinoma Cell Line

Glenn T. Merlino, Young-hua Xu, Nancy Richert, Adrian J. L. Clark, Shunsuke Ishii, Susan Banks-Schlegel, and Ira Pastan

Laboratory of Molecular Biology, Division of Cancer Biology and Diagnosis, and Laboratory of Human Carcinogenesis, Division of Cancer Etiology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Abstract

The human epidermal growth factor (EGF) receptor is known to be homologous to the *v-erb B* oncogene protein of the avian erythroblastosis virus. Overexpression of the EGF receptor gene in A431 epidermoid carcinoma cells is due to gene amplification. In this study, a variety of squamous cell carcinomas were examined and one, SCC-15, contained high levels of the EGF receptor as determined by immunoprecipitation via an EGF receptor-specific polyclonal antibody. Using a cloned EGF receptor complementary DNA as a probe, the level of EGF receptor RNA was found to be elevated four-fold in SCC-15 relative to normal cultured keratinocytes. When the same probe was used to identify EGF receptor gene fragments on a genomic DNA blot, the SCC-15 cell line was shown to possess an EGF receptor gene copy number amplified four to five times. Gene amplification results in the enhancement in the level of the EGF receptor in several carcinomas and could be responsible for the appearance of the transformed phenotype in these cells.

Introduction

The epidermal growth factor (EGF)¹ stimulates growth and elicits a wide variety of rapid and delayed responses by binding to high-affinity cell-surface receptors which are 170-kD glycoproteins (1). Recently, EGF receptor peptides have been sequenced and found to be homologous to the avian erythroblastosis virus *erb B* oncogene product (2), suggesting that the EGF receptor gene is the human *c-erb B* oncogene. A431 epidermoid carcinoma cells possess a very large number of EGF receptors (3), and the EGF receptor gene is amplified ~30-fold (4-6). This amplification is responsible for the overexpression of the EGF receptor protein in these cells (4-6).

A cell culture system has been developed permitting serial cultivation of keratinocytes, whose growth is modulated by EGF (7). Such methods have been used to establish cell lines from squamous cell carcinomas of the oral epithelium (8). Because of the role of EGF in keratinocyte development, we quantified EGF receptor protein and RNA in several squamous

cell carcinomas. One cell line, SCC-15, was found to contain high amounts of receptor protein and RNA, and a four- to fivefold amplification of the gene.

Methods

The squamous cell carcinomas established by Rheinwald and Beckett (8) were obtained from, and maintained according to the American Type Culture Collection (Rockville, MD). 1623 was originally designated as SCC-15; 1628 as SCC-25; and 1629 as SCC-9 (8). Normal human esophageal epithelial cells were grown as reported (9). Maintenance of other cell lines was as described elsewhere (10). Proteins were labeled with [³⁵S]methionine and immunoprecipitated as previously described (10). PolyA⁺ RNA was isolated by guanidine isothiocyanate solubilization and CsCl centrifugation (11), and oligo(dT)-affinity chromatography. RNA (Northern) blotting was performed as described (11, 12). High molecular weight DNA was isolated (4) and analyzed by DNA (Southern) blotting (4, 10, 13). The EGF receptor complementary DNA (cDNA) clone pE7 was constructed and isolated from an A431 cDNA library (11). DNA fragments were ³²P-labeled by nick translation.

Results

A large number of cell lines were initially screened for EGF receptor levels by determining their ability to be killed by an EGF-pseudomonas exotoxin conjugate, a technique described previously (10). Several squamous cell carcinomas were found to be relatively sensitive to the EGF-toxin conjugate, including SCC-25, SCC-9, and particularly SCC-15, all derived from the human tongue (8). These three cell lines were labeled with [³⁵S]methionine, and their extracts immunoprecipitated with a goat polyclonal antibody to the EGF receptor, affinity-purified as described (10). When compared with A431 cells, which make very large amounts of the EGF receptor, SCC-25 and SCC-9 make moderate amounts and SCC-15 high amounts of the receptor (Fig. 1, lane *a* vs. *e*, *g*, and *c*). Quantitation of the immunoprecipitation data revealed that SCC-15, SCC-25, and SCC-9 make 41, 15, and 4% of the amount of EGF receptor made by A431 cells, respectively.

Because SCC-15 cells had high levels of receptor, polyA⁺ RNA was isolated from these cells, electrophoretically fractionated on agarose, and analyzed by RNA (Northern) blotting. A cloned A431 cDNA (pE7) encoding the EGF receptor (11) was ³²P-labeled and used as a hybridization probe to visualize EGF receptor RNAs. Fig. 2 *A* shows that SCC-15 contains both the 10- and 5.6-kilobase species of EGF receptor RNA. The levels are approximately four- to fivefold higher than those found in either KB or A498 kidney carcinoma cells; these cell lines were previously found to possess readily de-

Received for publication 21 November 1984.

1. Abbreviations used in this paper: cDNA, complementary DNA; EGF, epidermal growth factor.

The Journal of Clinical Investigation, Inc.
Volume 75, March 1985, 1077-1079

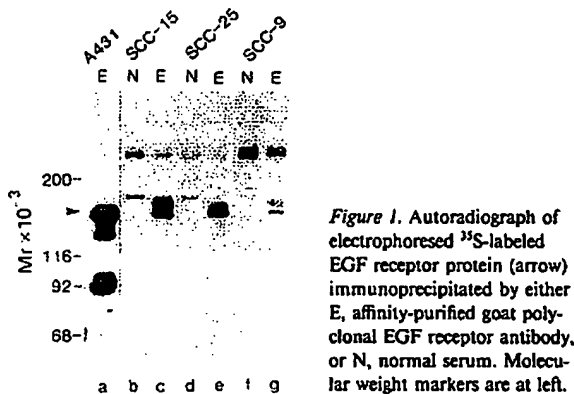


Figure 1. Autoradiograph of electrophoresed ^{35}S -labeled EGF receptor protein (arrow) immunoprecipitated by either E, affinity-purified goat polyclonal EGF receptor antibody, or N, normal serum. Molecular weight markers are at left.

etectable levels of both receptor RNAs (10). Fig. 2 *B* shows that cultured human epithelial cells contain EGF receptor-specific RNA (HEIA, lane *e*) whose levels are higher than an early passage human fibroblast D551 (lane *f*), equivalent to A498 (lane *g*), but much lower than SCC-15 (lane *b*).

To determine if an elevated gene copy number was associated with enhanced expression of the EGF receptor gene in SCC-15 cells, genomic DNA was isolated from normal cultured epithelial cells (HEIA) and SCC-15 cells, digested with *Hind*III, electrophoretically fractionated, and subjected to DNA blotting analysis. An EGF receptor cDNA (pE7) was used as a hybridization probe to identify receptor DNA fragments. Fig. 3 *A* reveals that the SCC-15 genome contains four- to fivefold amplified EGF receptor gene sequences relative to normal epithelial cells (lane *a* vs. *b*). Analysis of β -actin gene fragments on the same filter by hybridization to a chick actin cDNA probe indicated that equal amounts of DNA were loaded per well (data not shown). Digested SCC-15 DNA had to be diluted about fourfold (Fig. 3 *B*, lane *e*) to approximate the signal intensity of receptor DNA fragments from SCC-25, SCC-9, and KB cells (lanes *g*–*i*). The KB cell EGF receptor gene is known not to be amplified (10).

Discussion

It may be significant that A431 carcinoma cells are not unique in their possession of amplified EGF receptor genes. We report here that the EGF receptor gene in squamous cell carcinoma SCC-15 is amplified four- to fivefold relative to normal epithelial

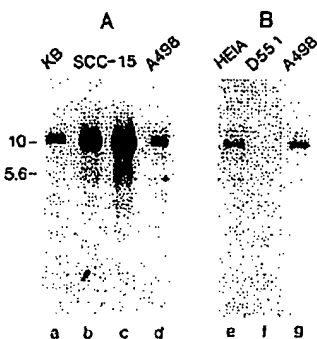


Figure 2. RNA blot analysis of polyA⁺ RNAs using the ^{32}P -labeled EGF receptor cDNA probe pE7. (*A*) and (*B*) are autoradiographs from two separate gels. Sizes are in kilobases (left). 5 (lanes *a*, *b*, *d*, and *g*) or 10 (lanes *c*, *e*, and *f*) μg of RNA were loaded.

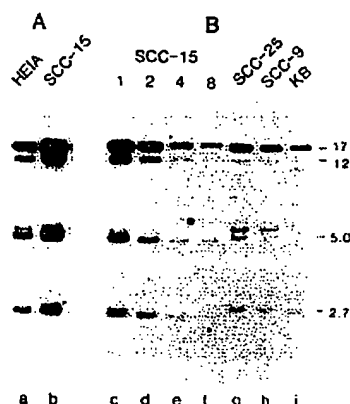


Figure 3. DNA blot analysis of *Hind*III-digested genomic DNAs using the pE7 probe (see Fig. 2). (*A*) and (*B*) are autoradiographs from two separate gels. Sizes are in kilobase pairs (right). 10 μg of DNA was loaded except in (*B*), lanes *d*–*f*, which represent serial dilutions of the 1623 DNA shown in lane *c*.

cells. The amplification of the EGF receptor gene may cause the initiation or maintenance of the malignant state in some human cells.

Previously, we reported that a variety of transformed cell lines synthesize relatively high amounts of both EGF receptor protein and messenger RNA (10). It is conceivable that a moderate or even a small increase in the level of the EGF receptor leads to a change in the cellular phenotype, as has been demonstrated for the *src* gene product (14). If this hypothesis is correct, then even a minor amplification of the EGF receptor gene copy number could contribute to the onset of tumorigenesis. Hendler and Ozanne (15) have examined lung squamous cell carcinomas and found that they contain a 2.5–5-fold increase in EGF receptor levels.

References

1. Savage, C. R., and S. J. Cohen. 1972. Epidermal growth factor and a new derivation: rapid isolation procedure and biological and chemical characterization. *J. Biol. Chem.* 247:7609–7611.
2. Downward, J., Y. Yarden, E. Mayes, G. Scarce, N. Totty, P. Stockwell, A. Ullrich, J. Schlessinger, and M. D. Waterfield. 1984. Close similarity of the epidermal growth factor receptor and v-erb B oncogene protein sequence. *Nature (Lond.)* 307:521–527.
3. Fabricant, R. N., J. E. DeLarco, and G. J. Todaro. 1977. Nerve growth factor receptor on human melanoma cells in culture. *Proc. Natl. Acad. Sci. USA* 74:565–569.
4. Merlino, G. T., Y.-H. Xu, S. Ishii, A. J. L. Clark, K. Semba, K. Toyoshima, T. Yamamoto, and I. Pastan. 1984. Amplification and enhanced expression of the epidermal growth factor receptor gene in A431 human carcinoma cells. *Science (Wash. DC)* 224:417–419.
5. Ullrich, A., L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. V. Mayes, N. Whittle, M. D. Waterfield, and P. H. Seeburg. 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermal carcinoma cells. *Nature (Lond.)* 309:418–425.
6. Lin, C. R., W. S. Chen, W. Kruger, L. S. Stolarsky, W. Weber, R. M. Evans, I. M. Verma, G. N. Gill, and M. G. Rosenfeld. 1984.

Expression cloning of human EGF receptor complementary DNA: gene amplification and three related messenger RNA products in A431 cells. *Science (Wash. DC)*. 224:843-848.

7. Rheinwald, J. G., and H. Green. 1975. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell*. 6:331-344.

8. Rheinwald, J. G., and M. A. Beckett. 1981. Tumorigenic keratinocyte lines requiring anchorage and fibroblast support cultured from human squamous cell carcinomas. *Cancer Res.* 41:1657-1663.

9. Banks-Schlegel, S. P., and C. C. Harris. 1983. Tissue-specific expression of keratin proteins in human esophageal and epidermal epithelium and their cultured keratinocytes. *Exp. Cell Res.* 146:271-280.

10. Xu, Y.-H., N. Richert, S. Ito, G. T. Merlino, and I. Pastan. 1984. Characterization of epidermal growth factor receptor gene expression in malignant and normal human cell lines. *Proc. Natl. Acad. Sci. USA*. 81:7308-7312.

11. Xu, Y.-H., J. Ishii, A. J. L. Clark, M. Sullivan, R. K. Wilson, D. P. Ma, B. A. Roe, G. T. Merlino, and I. Pastan. 1984. Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs overproduced in A431 carcinoma cells. *Nature (Lond.)*. 309: 806-810.

12. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA*. 77:5201-5205.

13. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.

14. Jakobovits, E. B., J. E. Majors, and H. E. Varmus. 1984. Hormonal regulation of the Rous sarcoma src gene via a heterologous promoter defines a threshold dose for cellular transformation. *Cell*. 38: 757-765.

15. Hendler, F. J., and B. W. Ozanne. 1984. Human squamous cell lung cancers express increased epidermal growth factor receptors. *J. Clin. Invest.* 74:647-651.

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft^{‡§}, Thomas Thykjaer^{||}, Frederic M. Waldman^{||}, Hans Wolf^{**}, and Jullo E. Celis^{‡‡}

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. *Molecular & Cellular Proteomics* 1:37-45, 2002.

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this

From the [‡]Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and [§]Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, ^{||}AROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, ^{||}UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ^{‡‡}Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of *c-erb-B2*, *cyclin d1*, *ems1*, and *N-myc* (3-5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of *c-myc* copy number increase was observed without concomitant *c-myc* protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)¹ have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q-, 11p-, 1q+, 11q13+, 17q+, and 20q+ (7-12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary).

¹ The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

Gene Copy Numbers, Transcripts, and Protein Levels

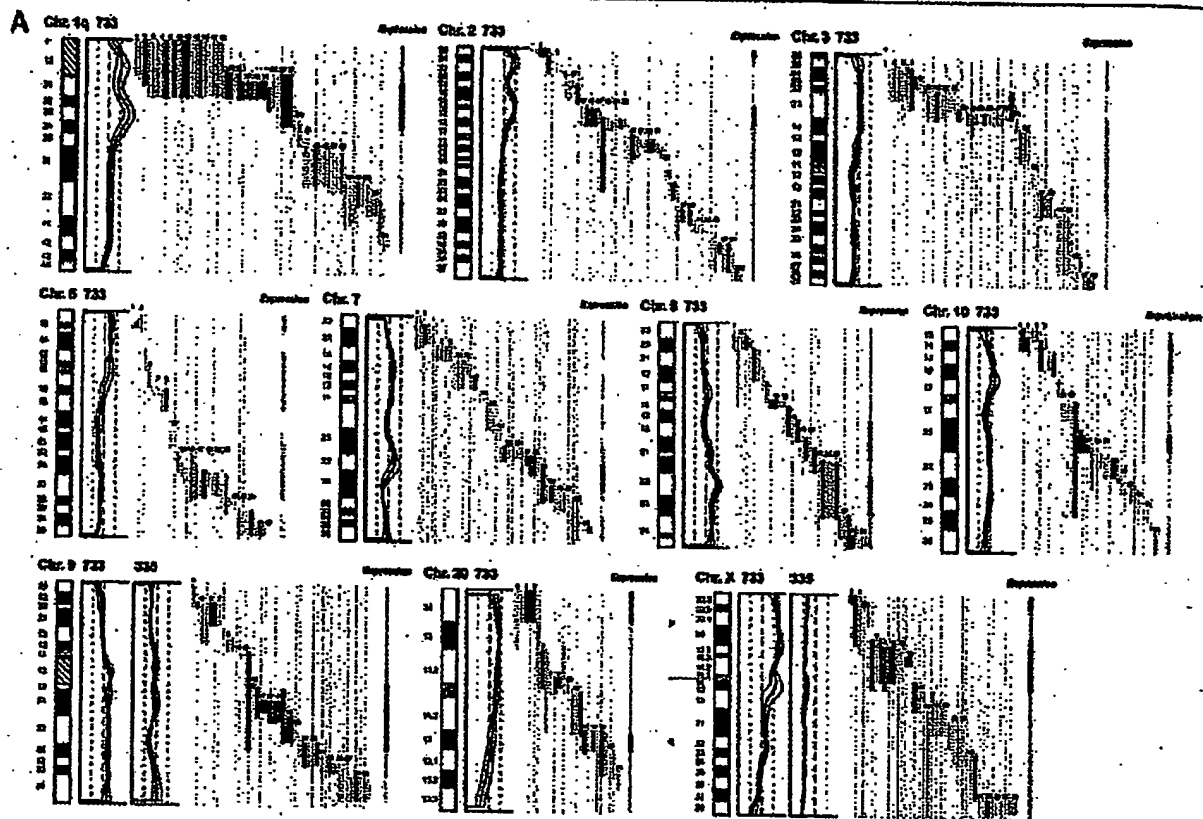


Fig. 1. DNA copy number and mRNA expression level. Shown from left to right are chromosome (Chr.), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. **A**, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. **B**, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (left). The bold curve in the ratio profile represents a ratio value of 1 (no change), and the vertical lines next to it (dotted) indicate a ratio of 0.5 (left) and 2.0 (right). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the right of the invasive tumor profile. The colored bars represent one gene each, identified by the running numbers above the bars (the name of the gene can be seen at www.MDL.DK/sdata.html). The bars indicate the purported location of the gene, and the colors indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart: >2-fold increase (black), >2-fold decrease (blue), no significant change (orange). The bar to the far right, entitled Expression shows the resulting change in expression along the chromosome; the colors indicate that at least half of the genes were up-regulated (black), at least half of the genes down-regulated (blue), or more than half of the genes are unchanged (orange). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at -80°C . Total RNA was isolated using the RNeasy 8 RNA isolation method (WAK-Chemie Medical GmbH). poly(A)⁺ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript[®] choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the MEGAscript[®] *in vitro* transcription kit (Ambion). Biotin-labeled CTP and

UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94°C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6 \times SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95°C for 5 min, subsequently cooled to 40°C , and loaded onto the Affymatrix probe array cartridge. The probe array was then incubated for 16 h at 40°C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6 \times SSPE-T at 25°C followed by 4 washes in 0.5 \times SSPE-T at 50°C . The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 $\mu\text{g}/\text{ml}$ (Molecular Probes) in 6 \times SSPE-T

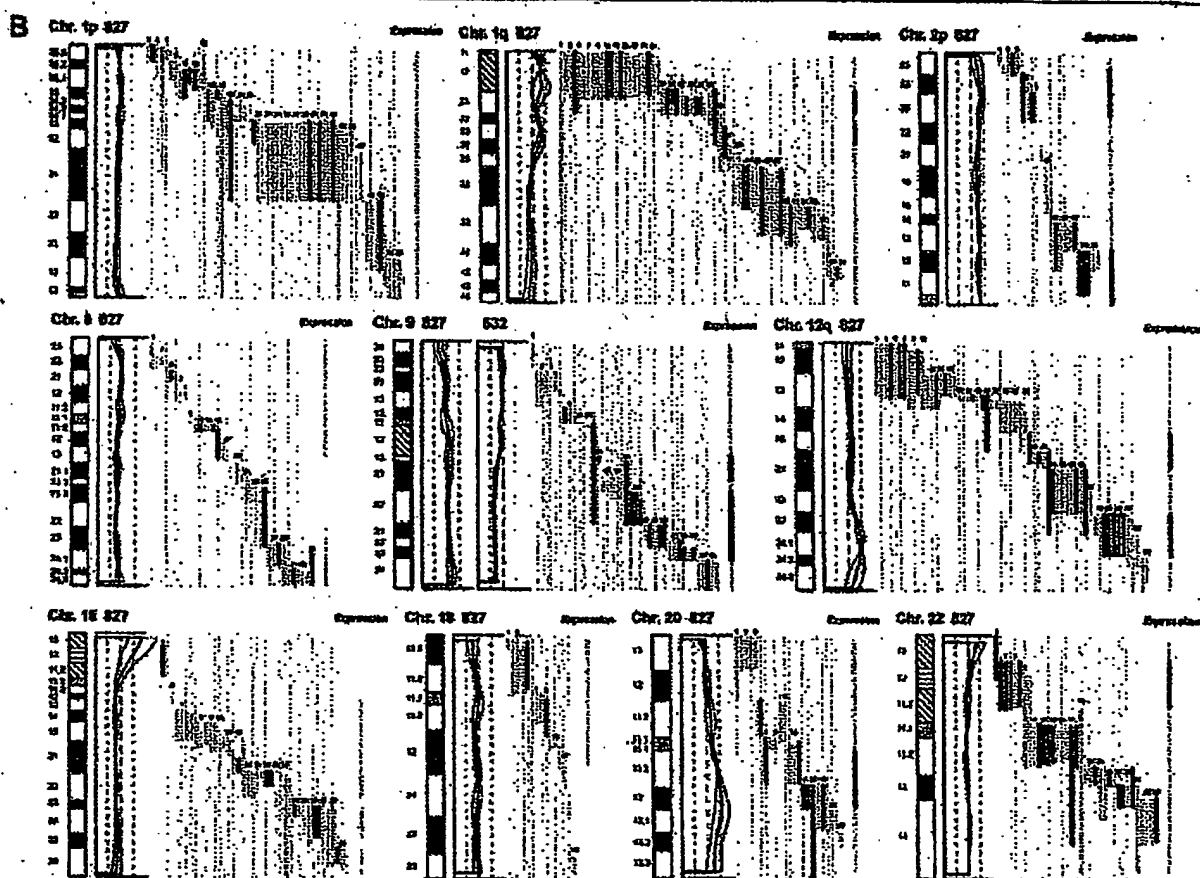


Fig. 1—continued

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μ l for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at -20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgi-bin/cells.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-

labeled reference DNA (200 ng), and human Cot-1 DNA (20 μ g) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μ g/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

Gene Copy Numbers, Transcripts, and Protein Levels

TABLE I
Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335		Concordance	CGH alterations	Tumor 827 vs. 532		Concordance
	Expression change clusters				Expression change clusters		
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%		10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%	
10 Loss	1 Up-regulation 6 Down-regulation 4 No change	50%		12 Loss	3 Up-regulation 2 Down-regulation 7 No change	17%	
Expression change clusters	Tumor 733 vs. 335		Concordance	Expression change clusters	Tumor 827 vs. 532		Concordance
	CGH alterations				CGH alterations		
16 Up-regulation	11 Gain 2 Loss 3 No change	69%		17 Up-regulation	10 Gain 5 Loss 2 No change	59%	
21 Down-regulation	1 Gain 8 Loss 12 No change	38%		9 Down-regulation	0 Gain 3 Loss 6 No change	33%	
15 No change	3 Gain 3 Loss 9 No change	60%		21 No change	1 Gain 3 Loss 17 No change	81%	

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22–q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22–24+, 2q14.1–qter–, 3q12–q13.3–, 6q12–q22–, 9q34+, 11q12–q13+, 17+, and 20q11.2–q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1A) and 20q12 in TCC 827 (Fig. 1B).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive *versus* the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger

than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21–q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, top). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, top). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, bottom). Furthermore, as a control we looked at areas with no alter-

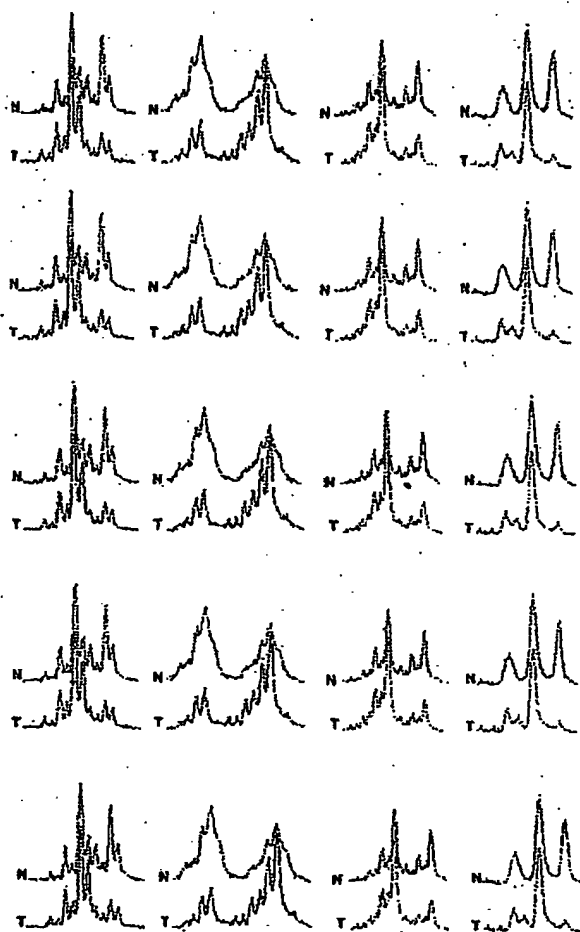


Fig. 3. Microsatellite analysis of loss of heterozygosity. Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (a) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (b) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (c) at chromosome 2p23 by D2S2251 close to general β -spectrin (gene number 11 on Fig. 1) and of (d) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The upper curves show the electropherogram obtained from normal DNA from leukocytes (N), and the lower curves show the electropherogram from tumor DNA (T). In all cases one allele is partially lost in the tumor amplicon.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Relation between Changes in mRNA and Protein Levels—2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH

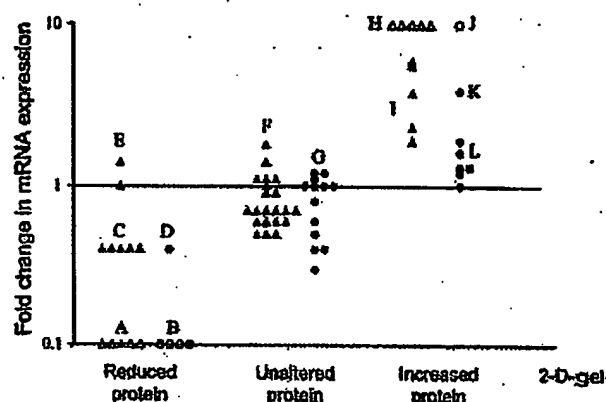


Fig. 4. Correlation between protein levels as judged by 2D-PAGE and transcript ratio. For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (horizontal axis). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (vertical axis). Δ , mRNAs that were scored as present in both tumors used for the ratio calculation; Δ , mRNAs that were scored as absent in the invasive tumors (along horizontal axis) or as absent in non-invasive reference (top of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (Δ) were scaled with background suppression, and TCCs 733 and 335 (\bullet) were scaled without suppression. Both comparisons showed highly significant ($p < 0.005$) differences in mRNA ratios between the groups. Proteins shown were as follows: Group A (from left), phosphoglucomutase 1, glutathione transferase class μ number 4, fatty acid-binding protein homologue, cytochrome 15, and cytochrome 13; B (from left), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytochrome 13, and calyculin; C (from left), α -enolase, hnRNP B1, 28-kDa heat shock protein, 14-3-3- ϵ , and pre-mRNA splicing factor; D, mesothelial keratin K7 (type II); E (from top), glutathione S-transferase- π and mesothelial keratin K7 (type II); F (from top and left), adenylate cyclase-associated protein, E-cadherin, keratin 19, calgizarin, phosphoglycerate mutase, annexin IV, cytoskeletal γ -actin, hnRNP A1, integral membrane protein calnexin (P90), hnRNP H, brain-type clathrin light chain-a, hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase β -1 subunit; G, (from top and left), TCP20, calgizarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytochrome 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphate dehydrogenase, glutathione S-transferase- π , and keratin 8; H (from left), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD+-dependent 15-hydroxyprostaglandin dehydrogenase; I (from top), prolyl 4-hydroxylase β -subunit, cytochrome 20, cytochrome 17, prohibitin, and fructose 1,6-bisphosphatase; J, annexin II; K, annexin IV; L (from top and left), 90-kDa heat shock protein, prolyl 4-hydroxylase β -subunit, α -enolase, GRP 78, cyclophilin, and cofilin.

gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

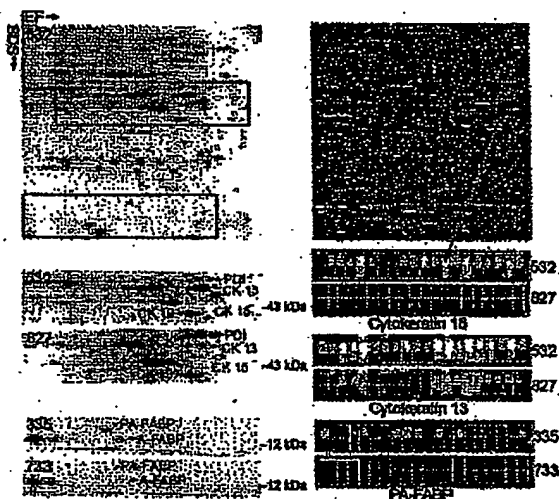


Fig. 5. Comparison of protein and transcript levels in invasive and non-invasive TCCs. The upper part of the figure shows a 2D gel (left) and the oligonucleotide array (right) of TCC 532. The red rectangles on the upper gel highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (red annotation). The tile on the array containing probes for cytokeratin 15 is enlarged below the array (red arrow) from TCC 532 and is compared with TCC 827. The upper row of squares in each tile corresponds to perfect match probes; the lower row corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as black, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (8151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the bottom of the figure (left) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the right we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). IEF, Isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a known chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FABP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II
Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration ^a	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres ^a	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA)-FABP	8q21.2	827	Loss	10-Fold down	Decrease
FABP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up ^b	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	6.7-/1.8-Fold up	Increase
hnRNPA1	7p15	827	Loss	2.5-Fold down	Decrease

^a Abs, absent; Pres, present.

^b In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (e.g. In TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; In TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17-19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p-, 9q-, 1q+, Y- (2, 6), and in pT1 tumors, 2q-, 11p-, 11q-, 1q+, 5p+, 8q+, 17q+, and 20q+ (2-4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p- and 9q22-q33- and 9q- and Y-, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22-24 amplification (seen in both tumors), 11q14-q22 loss, the latter often linked to 17q+ (both tumors), and 1q+ and 9p-, often linked to 20q+ and 11q13+ (both tumors) (7-9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal

arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicates that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (e.g. signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker et al. (26) in yeast.

(Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript.) One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

Acknowledgments—We thank Mlle Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Nikolaj Ørntoft, and Lynn Bjerke for technical help and Thomas Gíngeras, Christine Harrington, and Morten Østergaard for valuable discussions.

* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordisk, the Danish Biotechnology Program, the Frenkel Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk.

REFERENCES

- Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* 396, 643-649.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41-46.
- de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* 83, 167-161.
- Brugler, P. P., Tamini, Y., Shuaring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* 12, 1747-1753.
- Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* 50, 1459-1463.
- Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) *c-myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am. J. Pathol.* 146, 1131-1139.
- Richter, J., Jiang, F., Gorog, J. P., Sartorius, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* 57, 2860-2864.
- Richter, J., Belfa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* 153, 1615-1621.
- Bruch, J., Wöhr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Moller, P., Sauter, G., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* 23, 167-174.
- Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* 58, 3555-3560.
- Simon, R., Burger, H., Brinkschmidt, C., Bocker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* 185, 345-351.
- Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* 110, 87-93.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 15, 1359-1367.
- Christensen, M., Sunde, L., Bolund, L., and Ørntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* 59, 167-177.
- Celis, J. E., Østergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, L., Hejn, B., Wolf, H., Ørntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* 56, 4782-4790.
- Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) In *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222-230. Academic Press, Orlando, FL.
- Ohlsson, R., Tycko, B., and Sapiezka, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* 14, 435-438.
- Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the Interleukin-2 locus. *Science* 279, 2118-2121.
- Brannan, C. L., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* 9, 164-170.
- Ohlsson, R., Cui, H., He, L., Pfeifer, S., Maimitkumpa, H., Jiang, S., Feinberg, A. P., and Hedberg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* 59, 3888-3892.
- Cui, H., Hedberg, F., He, L., Nordenskjöld, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* 57, 4469-4473.
- Gafitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* 285, 251-254.
- Tsao, J., Yatabe, Y., Mark, I. D., Haiyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* 29, 28-32.
- Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10632-10638.
- Anderson, L., and Selinger, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533-537.
- Ideker, T., Thorsson, V., Raniish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 828-834.

Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer^{1,2}

Elizabeth Hyman,³ Päivikki Kauraniemi,³ Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Onti Mönni, Abdel Elkahloun, Olli-P. Kallioniemi, and Anne Kallioniemi⁴

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E.H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E.H., P.K., S.H., M.W., S.M., E.R., M.R., A.E., O.P., A.K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P.K., A.K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S.H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G.S.]; and Biomedicum Blochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O.M.]

ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.3% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1-6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over

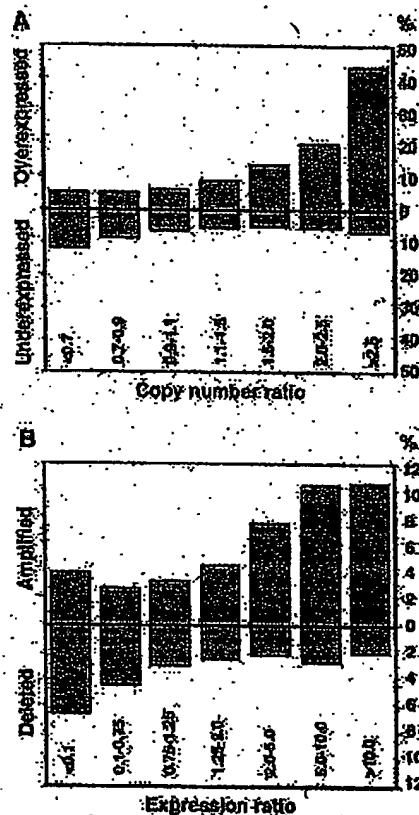


Fig. 1. Impact of gene copy number on global gene-expression levels. A, percentage of over- and underexpressed genes (Y axis) according to copy number ratio (X axis). Threshold values used for over- and underexpression were >2.184 (global upper 7% of the cDNA ratios) and <0.4826 (global lower 7% of the expression ratios). B, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were >1.5 and <0.7 .

20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH⁵ (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (a) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (b) identify and characterize those genes whose mRNA expres-

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

²Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org>).

³Contributed equally to this work.

⁴To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lemminkäisenkatu 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

⁵The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence *in situ* hybridization; RT-PCR, reverse transcription-PCR.

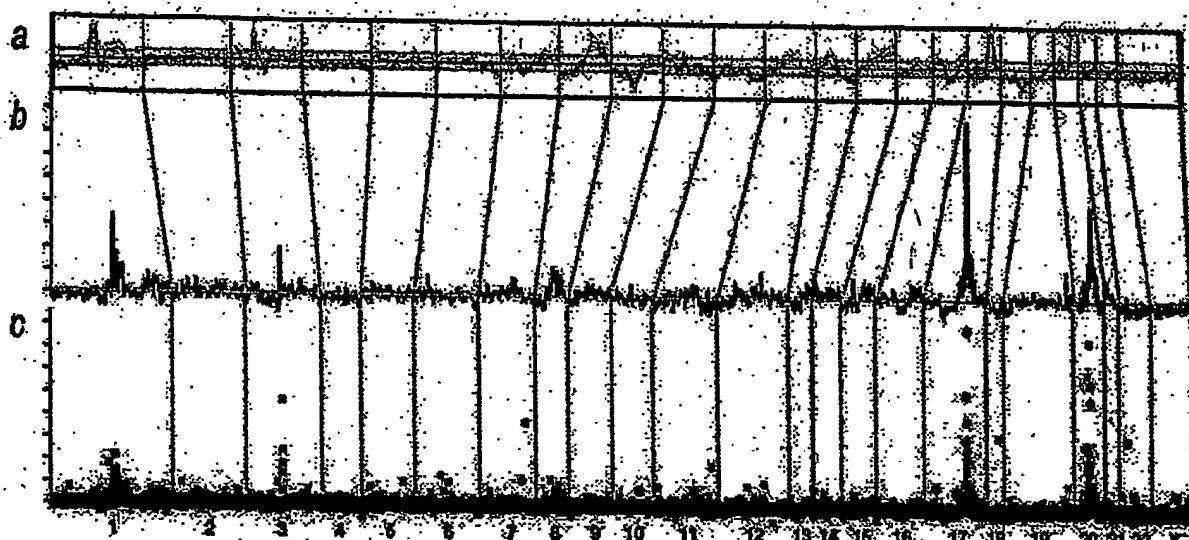


Fig. 2. Genomic-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (blue line) across the entire genome from 1p telomere to Xq telomere is shown along with ± 1 SD (orange lines). The black horizontal line indicates a ratio of 1.0; red line, a ratio of 0.8; and green line, a ratio of 1.2. *B–C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. Red horizontal line, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The bright green dots indicate the upper 2%, and dark green dots, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); bright green dots indicate the lowest 2%, and dark green dots, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with black crosses. The chromosomes numbers are shown at the bottom of the figure, and chromosome boundaries are indicated with a dashed line.

sion is most significantly associated with amplification of the corresponding genomic template.

MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μ g of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *AluI* and *RsaI* (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μ g of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μ g of reference RNA were labeled with Cy3-dUTP and 3.5 μ g of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units)

were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, w_g , for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1} , σ_{g1} , and m_{g0} , σ_{g0} denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by α . A low α (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.⁶ A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.⁷ The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

⁶ Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cdna.html.

⁷ Internet address: www.genome.ucsc.edu.

Table 1. Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1-7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11-8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3-8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22-q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12-q21.2	39.79	42.80	3.0
17q21.32-q21.33	52.47	55.80	3.3
17q22-q23.3	63.81	69.70	5.9
17q23.3-q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12-q13.13	46.45	49.43	3.0
20q13.2-q13.32	51.32	59.12	7.8

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

FISH. Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

RT-PCR. The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5'-GAGCAGAGGACTCGGACTT-3' and 5'-GCGTCAGGTAGCGATTGTAG-3'.

RESULTS

Global Effect of Copy Number on Gene Expression. 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (i.e., belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1B). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

Identification of Distinct Breast Cancer Amplicons. Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal

CGH were validated, with 1q21, 17q12-q21.2, 17q22-q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65-39.25 Mb), and 17q21.3 (52.47-55.80 Mb).

Direct Identification of Putative Amplification Target Genes. The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2C shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22-q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11-p12 (Fig. 3A). In BT-474, the two known amplicons at 17q12 and 17q22-q23 contained numerous highly overexpressed genes (Fig. 3B). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3B, inset) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

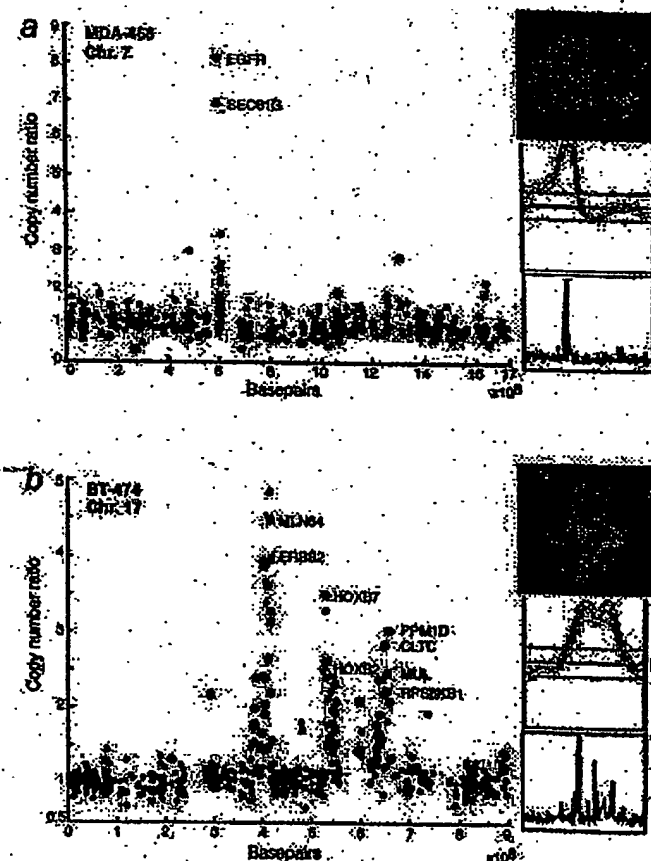


Fig. 3. Annotation of gene expression data on CGH microarray profiles. A, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (red dots) and include the *EGFR* oncogene. B, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (red) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2C. Insets show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (red) and chromosome 7 centromere probe (green) to MDA-468 (A) and *HOXB7*-specific probe (red) and chromosome 17 centromere (green) to BT-474 cells (B).

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22-24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1-2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29-32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein S6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (a) evidence of a prominent global influence of copy number changes on gene expression levels; (b) a high-resolution map of 24 independent amplicons in breast cancer; and (c) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association

between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

REFERENCES

- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* (Wash. DC), 286: 531-537, 1999.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Traub, T., Yu, X., et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* (Lond.), 403: 503-511, 2000.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, B., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* (Lond.), 406: 536-540, 2000.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnson, H., Akslen, L. A., et al. Molecular portraits of human breast tumours. *Nature* (Lond.), 406: 747-752, 2000.
- Dhanasekaran, S. M., Barrette, T. R., Ghossein, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. *Nature* (Lond.), 412: 822-826, 2001.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. Gene expression patterns of breast carcinomas distinguish tumor subtypes with clinical implications. *Proc. Natl. Acad. Sci. USA*, 98: 10869-10874, 2001.
- Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. *Semin. Cancer Biol.*, 9: 125-138, 1999.
- Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. *J. Clin. Oncol.*, 19: 32-40, 2001.
- Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larremans, M. L., Tapper, J., Petro, H., El-Rifai, W., et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107-1123, 1998.
- Knuutila, S., Autio, K., and Aalto, Y. Online access to CGH data of DNA sequence copy number changes. *Am. J. Pathol.*, 157: 689, 2000.
- DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.*, 14: 457-460, 1996.
- Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.*, 6: 639-645, 1996.
- Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Buzovnic, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. In: F. J. Livesey and S. P. Hunt (eds.), *Functional Genomics*, pp. 113-137. Oxford: Oxford University Press, 2000.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, 23: 361-46, 1999.
- Monni, O., Birtlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA*, 98: 5711-5716, 2001.
- Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics*, 2: 364-374, 1997.
- Birtlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Thorntorn, J., Haas, P., Bucher, C., Sauter, G., et al. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.*, 92: 1252-1259, 2000.
- Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. *Cytometry*, 45: 83-86, 2001.
- Kauraniemi, P., Birtlund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.*, 61: 8235-8240, 2001.
- Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Mallard, K., Giddings, L., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. *Genes Chromosomes Cancer*, 34: 104-114, 2002.
- Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer Res.*, 62: 2625-2629, 2002.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

- Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.*, 25: 333-337, 2000.
23. Virtanen, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc. Natl. Acad. Sci. USA*, 98: 1124-1129, 2001.
24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Puzello, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linehan, W. M., Cima, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res.*, 61: 8143-8149, 2001.
25. Bärndt, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22-q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer*, 20: 372-376, 1997.
26. Tamer, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi, O. P. Independent amplification and frequent co-amplification of three nonsynthetic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res.*, 56: 3441-3445, 1996.
27. Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. *Exp. Cell Res.*, 248: 1-9, 1999.
28. Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. *J. Cell. Physiol.*, 188: 161-169, 2001.
29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol. Cell. Biol.*, 16: 4842-4851, 1996.
30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the HOXB7 gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. *Oncogene*, 16: 3285-3289, 1998.
31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. *Cancer Res.*, 61: 6532-6539, 2001.
32. Naor, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. *Proc. Natl. Acad. Sci. USA*, 98: 4060-4065, 2001.

Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack^{*†‡}, Therese Sørlie[§], Charles M. Perou[¶], Christian A. Rees^{||}, Stefanie S. Jeffrey^{**}, Per E. Lønning^{**}, Robert Tibshirani^{§§}, David Botstein^{¶¶}, Anne-Lise Børresen-Dale[§], and Patrick O. Brown^{††§§}

Departments of ^{*}Pathology, Genetics, [†]Surgery, [‡]Health Research and Policy, and ^{§§}Biochemistry, and [¶]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; [§]Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; ^{¶¶}Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and ^{||}Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., FGFR1 (8p11), MYC (8q24), CCND1 (11q13), ERBB2 (17q12), and ZNF217 (20q13)] and tumor suppressor genes [RB1 (13q14) and TP53 (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From

this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

Materials and Methods

Tumors and Cell Lines. Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org. Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

DNA Labeling and Microarray Hybridizations. Genomic DNA labeling and hybridizations were performed essentially as described in Pollack *et al.* (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. "Test" DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The "reference" (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

Data Analysis and Map Positions. Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at <http://rana.jbl.gov>). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by

Abbreviation: CGH, comparative genomic hybridization.

To whom reprint requests should be addressed at: Department of Pathology, Stanford University School of Medicine, CCSR Building, Room 3245A, 269 Campus Drive, Stanford, CA 94305-5176. E-mail: pollack1@stanford.edu.

**Present address: Zyomyx Inc., Hayward, CA 94545.

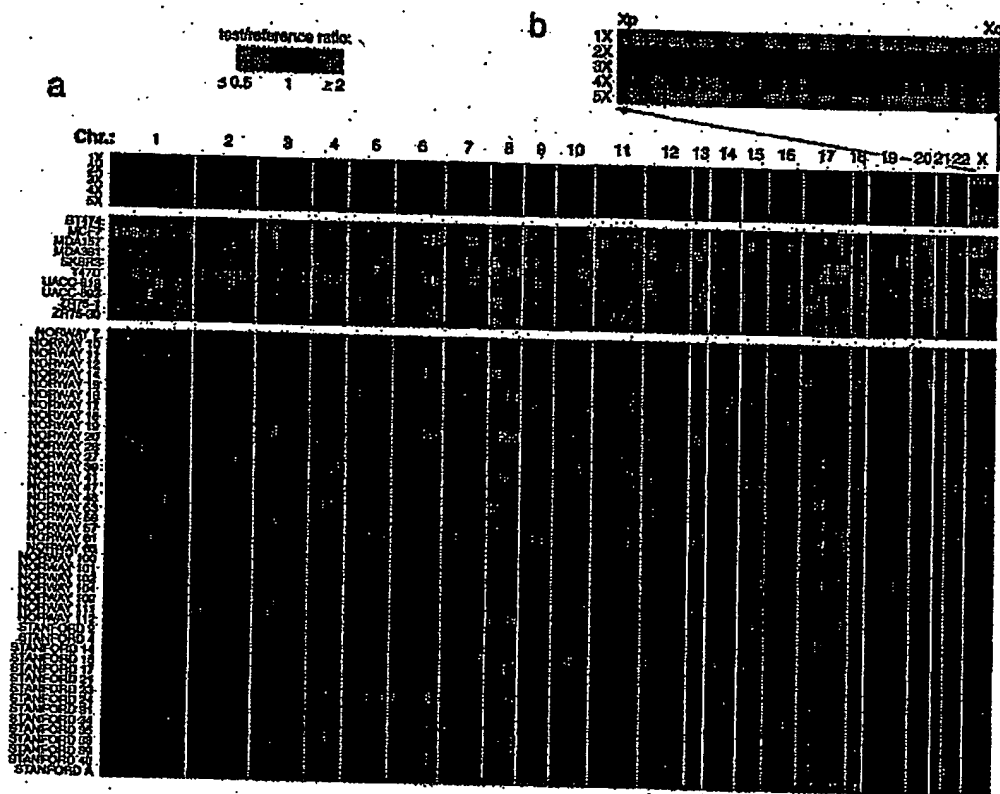


Fig. 1. Genome-wide measurement of DNA copy number alteration by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a log₂-based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (b) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (<http://genome.ucsc.edu/>; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1a; also see *Materials and Methods* for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (<http://genome.ucsc.edu/>) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and

deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1b), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47, XXX), 2- (48, XXXX), or 2.5-fold (49, XXXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1a), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2-4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total

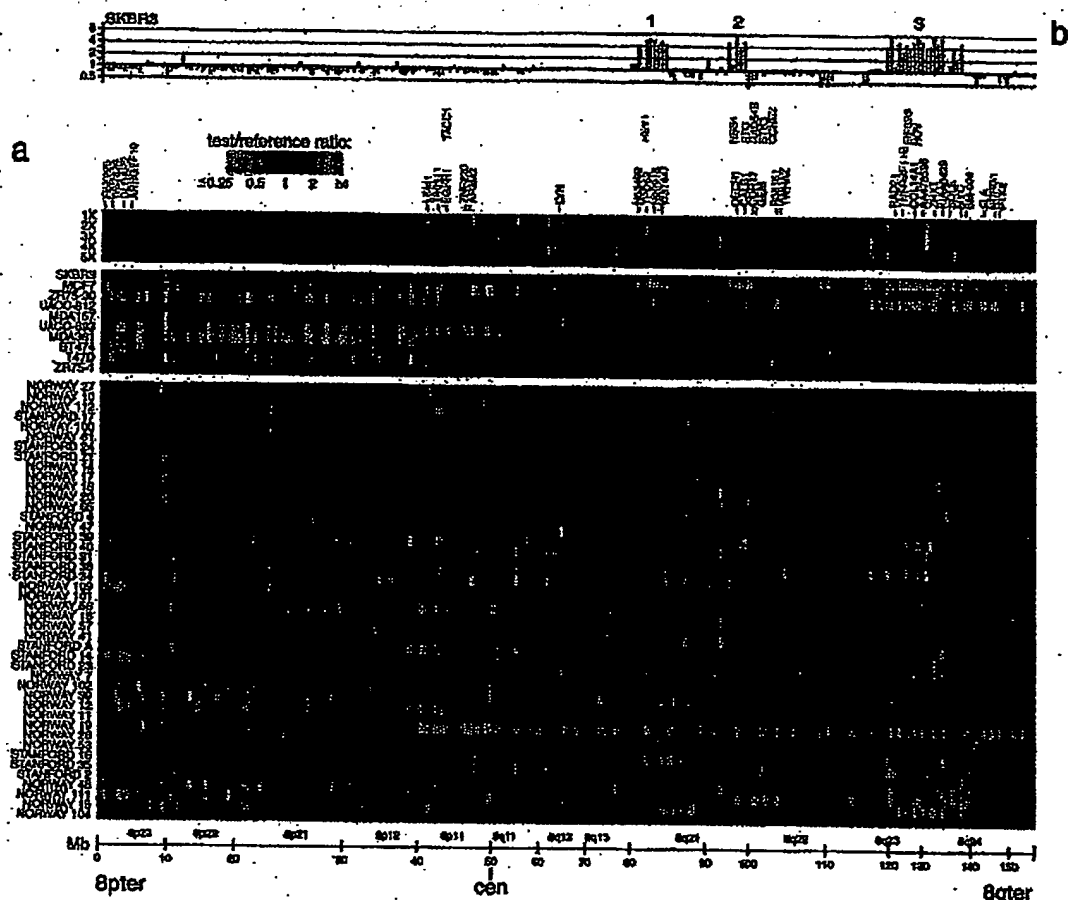


Fig. 2. DNA copy number alteration across chromosome 8 by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a log₂ pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (b) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a log₂ scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ($P = 0.008$), consistent with published CGH data (3), estrogen receptor negative ($P = 0.04$), and harboring TP53 mutations ($P = 0.0006$) (see Table 4, which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2a). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1–3 in Fig. 2b). For each of these regions we can define the

boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations

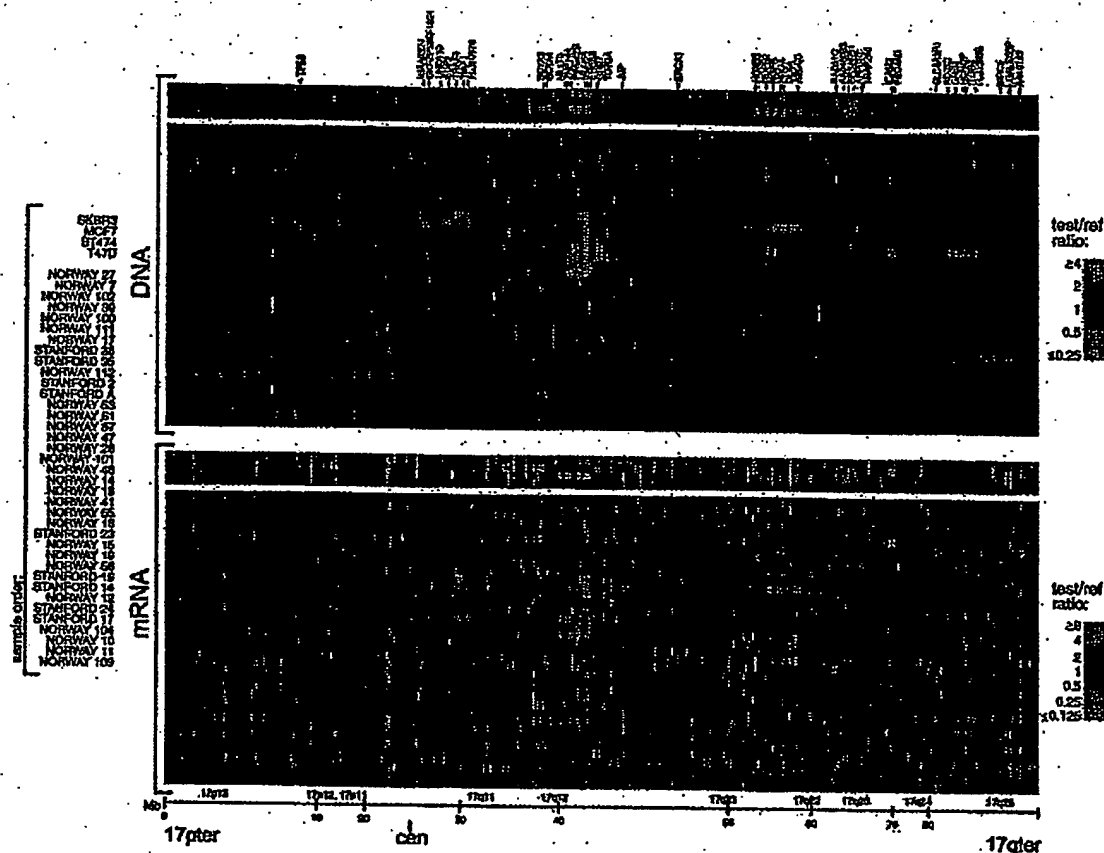


Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (Upper) and mRNA levels (Lower) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (Upper), and the identical sample order is maintained (Lower). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate \log_2 pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4 , and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4a). For both the

breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion (P values for pair-wise Student's t tests comparing adjacent classes: cell lines, 4×10^{-49} , 1×10^{-49} , 5×10^{-5} , 1×10^{-2} ; tumors, 1×10^{-43} , 1×10^{-214} , 5×10^{-41} , 1×10^{-4}). A linear regression of the average $\log(\text{DNA copy number})$, for each class, against average $\log(\text{mRNA level})$ demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4a, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4b). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4c), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4b) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37

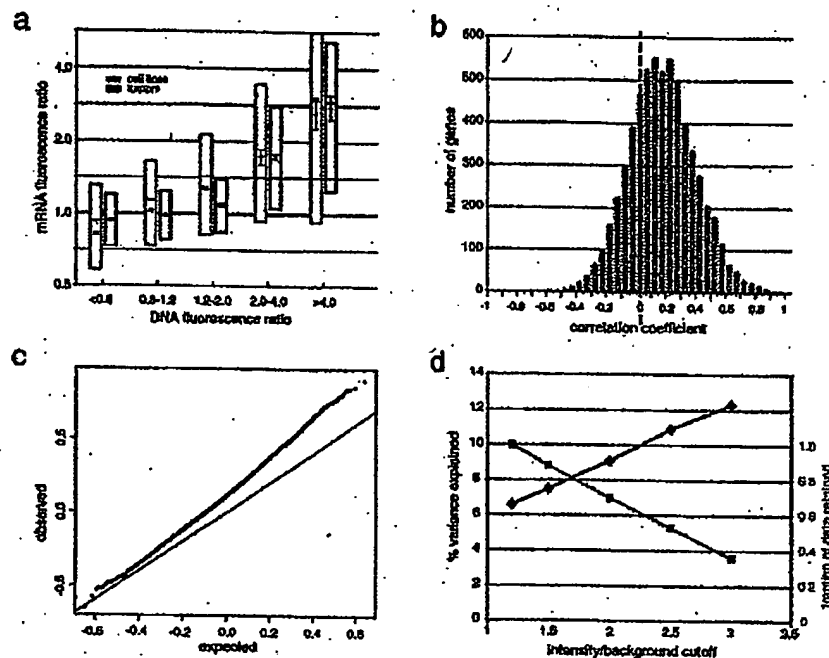


Fig. 4. Genome-wide influence of DNA copy number alterations on mRNA levels. (a) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio (log₂ scale) quartiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; Y-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. P values for pair-wise Student's t tests, comparing averages between adjacent classes (moving left to right), are 4×10^{-49} , 1×10^{-49} , 5×10^{-5} , 1×10^{-2} (cell lines), and 1×10^{-43} , 1×10^{-214} , 5×10^{-41} , 1×10^{-4} (tumors). (b) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (c) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (d) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4d). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background > 3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4d). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer

cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the

studies. For example, the study of Platzer *et al.* (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousand residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing

the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stoichiometric relationships in cell metabolism and physiology (e.g., proteasome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

We thank the many members of the P.O.B. and D.B. labs for helpful discussions. J.R.P. was a Howard Hughes Medical Institute Physician Postdoctoral Fellow during a portion of this work. P.O.B. is a Howard Hughes Medical Institute Associate Investigator. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, the Norwegian Cancer Society, and the Norwegian Research Council.

- Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) *Science* 258, 818-821.
- Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2156-2160.
- Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) *Genes Chromosomes Cancer* 21, 177-184.
- Forozan, F., Mahlamaki, E. H., Morni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) *Cancer Res.* 60, 4519-4525.
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nicklisch, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) *Genes Chromosomes Cancer* 18, 399-407.
- Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, L., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* 20, 207-211.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* 23, 41-46.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature (London)* 406, 747-752.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* 98, 10869-10874.
- Schuler, G. D. (1997) *J. Mol. Med.* 75, 694-698.
- Lander, E. S., Lipton, L. M., Birren, B., Nussbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature (London)* 409, 860-921.
- Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) *Genes Chromosomes Cancer* 22, 105-113.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., *et al.* (2000) *Nat. Genet.* 25, 333-337.
- Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., *et al.* (2001) *Cancer Res.* 61, 8143-8149.
- Platzer, F., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Wilson, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) *Cancer Res.* 62, 1134-1138.
- Albertson, D. G., Ylstra, B., Seagraves, R., Collins, C., Daike, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) *Nat. Genet.* 25, 144-146.
- Li, B., Yerganian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) *Proc. Natl. Acad. Sci. USA* 94, 14506-14511.
- Rasnick, D. & Duesberg, P. H. (1999) *Biochem. J.* 348, 621-630.

Research article

Open Access

Cyclin A and cyclin D1 as significant prognostic markers in colorectal cancer patients

Abeer A Bahnassy*¹, Abdel-Rahman N Zekri², Soumaya El-Houssini¹, Amal MR El-Shehaby³, Moustafa Raafat Mahmoud¹, Samira Abdallah⁴ and Mostafa El-Serafi⁵

Address: ¹Pathology Department, National Cancer Institute, Cairo University, Cairo, Egypt, ²Virology and Immunology Unit, Cancer Biology Department, National Cancer Institute, Cairo University, Cairo, Egypt, ³Biochemistry Department, Kasr El-Eini School of Medicine, Cairo University, Cairo, Egypt, ⁴Pathology Department, Kasr El-Eini School of Medicine, Cairo University, Cairo, Egypt and ⁵Medical Oncology Department, National Cancer Institute, Cairo University, Cairo, Egypt

Email: Abeer A Bahnassy* - chaya2000@hotmail.com; Abdel-Rahman N Zekri - ncizakri@starnet.com.eg; Soumaya El-Houssini - chaya2000@hotmail.com; Amal MR El-Shehaby - chaya2000@hotmail.com; Moustafa Raafat Mahmoud - ncizakri@starnet.com.eg; Samira Abdallah - chaya2000@hotmail.com; Mostafa El-Serafi - melserafi@starnet.com.eg

* Corresponding author

Published: 23 September 2004

Received: 25 April 2004

BMC Gastroenterology 2004, 4:22 doi:10.1186/1471-230X-4-22

Accepted: 23 September 2004

This article is available from: <http://www.biomedcentral.com/1471-230X/4/22>

© 2004 Bahnassy et al; licensee BioMed Central Ltd.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Colorectal cancer is a common cancer all over the world. Aberrations in the cell cycle checkpoints have been shown to be of prognostic significance in colorectal cancer.

Methods: The expression of *cyclin D1*, *cyclin A*, *histone H3* and *Ki-67* was examined in 60 colorectal cancer cases for co-regulation and impact on overall survival using immunohistochemistry, southern blot and in situ hybridization techniques. Immunoreactivity was evaluated semi quantitatively by determining the staining index of the studied proteins.

Results: There was a significant correlation between *cyclin D1* gene amplification and protein overexpression (concordance = 63.6%) and between *Ki-67* and the other studied proteins. The staining index for *Ki-67*, *cyclin A* and *D1* was higher in large, poorly differentiated tumors. The staining index of *cyclin D1* was significantly higher in cases with deeply invasive tumors and nodal metastasis. Overexpression of *cyclin A* and *D1* and amplification of *cyclin D1* were associated with reduced overall survival. Multivariate analysis shows that *cyclin D1* and *A* are two independent prognostic factors in colorectal cancer patients.

Conclusions: Loss of cell cycle checkpoints control is common in colorectal cancer. *Cyclin A* and *D1* are superior independent indicators of poor prognosis in colorectal cancer patients. Therefore, they may help in predicting the clinical outcome of those patients on an individual basis and could be considered important therapeutic targets.

Background

Colorectal cancer (CRC) is the third most common cancer in Western countries [1]. In Egypt, CRC has unique char-

acteristics that differ from that reported in other countries of the western society. It was estimated that 35.6% of the Egyptian CRC cases are below 40 years of age and patients

usually present with advanced stage, high grade tumors that carry more mutations [2]. This uniquely high proportion of early-onset CRC, the early and continuous exposure to hazardous environmental agents, the different mutational spectrum and the prevalent consanguinity in Egypt justify further studies [3]. It was proved that most cancers result from accumulation of genetic alterations involving certain groups of genes, the majority of which are cell cycle regulators that either stimulate or inhibit cell cycle progression [1]. Cell proliferation allows orderly progression through the cell cycle, which is governed by a number of proteins including *cyclins* and *cyclin* dependent kinases [4,5]. The *cyclins* belong to a superfamily of genes whose products complex with various *cyclin*-dependent kinases (*cdks*) to regulate transitions through key checkpoints of the cell cycle [6]. Abnormalities of several *cyclins* have been reported in different tumor types, implicating, in particular, *cyclin A*, *cyclin E* and *cyclin D* [6,7].

Cyclin D1 is a G1 *cyclin* that regulates the transition from G1 to S phase since its peak level and maximum activity are reached during the G1 phase of the cell cycle. Whereas

cyclin A is regarded a regulator of the transition to mitosis since it reaches its maximum level during the S and G2 phases [8]. The mechanisms likely to activate the oncogenic properties of the *cyclins* include chromosomal translocations, gene amplification and aberrant protein overexpression [7,9].

Several studies have shown that, *histone H3* mRNA expression can be used to identify the S phase fraction (SPF) through the in situ hybridization (ISH) technique [10,11]. The level of *histone H3* mRNA reaches its peak during the S phase and then drops rapidly at the G2 phase [12].

In face of the increasing incidence of CRC and its peculiar pattern in the Egyptian population, the present study was conducted to assess the role of Ki-67 (pan-cell cycle marker), *cyclin D1* (G1 phase marker), *histone H3* mRNA (S phase marker), *cyclin A* (S to G2 phase marker) in CRC. The expression level of these markers was correlated to the clinicopathologic features and the overall survival of patients.

Table 1: Clinicopathological features of patients in relation to the staining index (SI) of Ki-67, *cyclin D1*, *cyclin A*, *histone H3*

Variables	No. of cases	SI (mean + SD)			
		Ki-67	Cyclin D1	Cyclin A	Histone H3
Sex					
Male	36	18.0 ± 6.4	6.7 ± 4.3	12.7 ± 5.7	10.7 ± 5.3
Female	24	20.1 ± 5.8	8.8 ± 8.4	10.0 ± 6.0	10.7 ± 5.4
Age (years)					
≥50	41	11.7 ± 6.0*	5.6 ± 5.2	10.0 ± 5.3	6.0 ± 5.0*
<50	19	23.8 ± 5.6	7.7 ± 6.8	13.6 ± 5.7	22.0 ± 5.2
Tumor size (cm)					
<5.0	33	12.2 ± 6.3*	5.3 ± 3.8*	11.5 ± 6.1*	10.3 ± 4.9*
≥5.0	27	30.1 ± 6.2	22.8 ± 7.2	28.6 ± 5.6	24.0 ± 5.6
Histology					
Normal	20	3.5 ± 2.0*	0.6 ± 0.2*	2.3 ± 1.1*	2.2 ± 0.9
Carcinoma	60	30.3 ± 6.2	24.9 ± 6.3	27.2 ± 5.8	10.7 ± 5.3
GI	15	11.7 ± 6.2	6.6 ± 4.0	10.0 ± 5.4	11.4 ± 4.9
GII	21	11.8 ± 5.6	8.9 ± 3.6	12.3 ± 6.5	7.8 ± 5.4
GIII	24	30.0 ± 4.3	22.0 ± 8.1	27.0 ± 4.9	11.5 ± 5.4
Lymph node					
Negative	33	19.5 ± 7.0	5.4 ± 5.3*	11.9 ± 6.5	12.3 ± 5.5
Positive	27	21.3 ± 4.9	20.6 ± 6.9	12.5 ± 5.0	14.2 ± 5.0
Depth of Invasion					
m, sm	17	20.7 ± 6.7	3.1 ± 3.1*	11.9 ± 7.2	10.4 ± 5.1
beyond sm	43	21.9 ± 6.2	12.4 ± 6.5	12.2 ± 5.6	10.7 ± 5.4
Stage					
I	6	20.6 ± 6.7	5.7 ± 6.9	24.2 ± 6.9	11.1 ± 5.3
II	27	20.8 ± 6.9	5.3 ± 4.3	24.6 ± 6.0	10.4 ± 5.7
III	12	22.0 ± 5.4	7.7 ± 6.0	27.1 ± 5.2	10.4 ± 4.9
IV	15	24.7 ± 6.1	11.3 ± 9.6	27.5 ± 5.5	12.3 ± 6.2

* p. value < 0.05 (significant)

Methods

Tissue samples

Paraffin-embedded tumor tissues were obtained from 60 CRC patients (47 colon and 13 rectal carcinomas) that were diagnosed and treated at the National Cancer Institute, Cairo, Egypt during the period from January, 1997 to June, 2002. Clinicopathological data of the studied cases are illustrated in table 1. None of the patients received any chemotherapy or irradiation prior to surgery. Histological diagnosis of all cases was done by 2 independent pathologists according to the WHO Histological Classification. Tumors were staged according to the TNM staging system [13]. The depth of tumor invasion was classified as invasion of the mucosa including muscularis mucosa (m), invasion of the submucosa (sm), or invasion beyond the submucosa [8]. Normal colonic tissues were obtained from autopsy specimens (n = 20) and were used as a control. The actual survival rate of the patients was calculated from the date of resection to the date of death.

Immunohistochemistry

Four micron sections of each normal and tumor specimen were cut onto positive-charged slides; air dried overnight, de-paraffinized in xylene, hydrated through a series of graded alcohol and washed in distilled water and 0.01 PBS (pH 7.4). Slides were then processed for IHC as described by Handa et al. [8], using the following antibodies: Ki-67 (MIB-1, Dako), *cyclin A* (6E6; Novocastra, Newcastle-Upon-Tyne, UK) and *cyclin D1* (DCS-6, Dako). A case of invasive breast cancer was used as a positive control for Ki-67 and *cyclin A* whereas a case of mantle cell lymphoma was used as a control for *cyclin D1*. Negative controls were obtained by replacing the primary antibody by non-immunized rabbit or mouse serum.

Brown nuclear staining was regarded as a positive result for all studied markers. The proportion of positively-stained cells and the intensity of staining were scored in tumor and normal colorectal mucosal sections at medium power ($\times 200$). The degree of positive tumor staining (percentage of positive tumor cells in the examined section) was scored from 1–6 and the staining intensity was scored from 0–6 according to the pattern of staining in the examined section. Staining index (SI) was calculated by multiplying the cellularity and staining scores as described by King et al. [14].

In situ hybridization

All tumor samples and 5 normal controls were assessed for *histone H3* mRNA by ISH using the commercially available 550 base fluorescein-labeled DNA probe (Dako, Carpinteria, CA) as described by Nagao et al., 1996. This probe hybridizes to the whole mRNA transcript of the human *histone H3* gene including the 5' and 3' untranslated regions. Scoring of *histone H3* mRNA was performed

as for immunohistochemistry, however, hybridization signals were detected in the cytoplasm.

Molecular detection of cyclin D1 gene amplification

High molecular weight DNA was extracted from paraffin-embedded tissues of the tumor and normal colorectal mucosal samples as previously described [15]. The proportion of neoplastic and normal cells was determined in each tumor sample by examining hematoxylin and eosin-stained slides obtained from the edge of the specimen used for DNA extraction. Tumor samples were evaluated for amplification of *cyclin D1* if more than 75% of the examined sections were formed of neoplastic cells. Accordingly, 50 cases were eligible for the analysis. Ten micrograms of the extracted DNA was digested with *EcoRI*. DNA from selected cases was also digested with *BglII* and *HindIII*. Samples were separated on 0.8% agarose gels and transferred to Hybond-N membranes (Amersham Int., Amersham, UK). The membranes were hybridized with 50% formamide, $5 \times$ SSC, $5 \times$ Denhardt's, 500 μ g/ml denatured salmon sperm DNA, 10% dextran sulphate and 10^6 cpm/ml of 32 P-labeled *PRAD-1* probe for 24 h. Membranes were washed with $2 \times$ SSC, 0.1% SDS at room temperature for 30 min followed by $2 \times$ SSC, 0.1% SDS at 60°C for 30 min and $0.1 \times$ SSC, 0.1% SDS at 60°C for 1 h. Filters were autoradiographed using an intensifying screen at -70°C for 24–72 h. After being stripped free of the *PRAD-1* probe, the same blots were hybridized with 32 P-labeled *B-actin* probe to normalize against possible variations in the loading or transfer of DNA. The autoradiograms were analyzed using a densitometer. Intensities of *PRAD-1/cyclin D1* were normalized to the β -actin control bands. The degree of amplification was calculated from these normalized values. Amplification was considered when the signal of the tumor band was ≥ 2 -fold the value of the matched normal mucosa [16].

Statistical analysis

The Mann-Whitney non-parametric test was used to compare the SIs of pairs of subjects whereas the Kruskal-wallis was used for categorical data. Correlation between indices was performed using a simple linear regression test. The Kaplan-Meier method was used to create survival curves which were analyzed by the log-rank test. The impact of different variables on survival was determined using the Cox proportional hazards model. *p* values less than 0.05 were considered significant.

Results

The results of IHC are illustrated in figures 1 and 2. In general, the staining index (SIs) of all studied markers was higher in carcinomas than in normal colonic mucosal samples ($p = 0.0001$). Normal colorectal mucosa revealed positive immunostaining for Ki-67 in the lower half of the crypts only. A heterogeneous staining pattern was

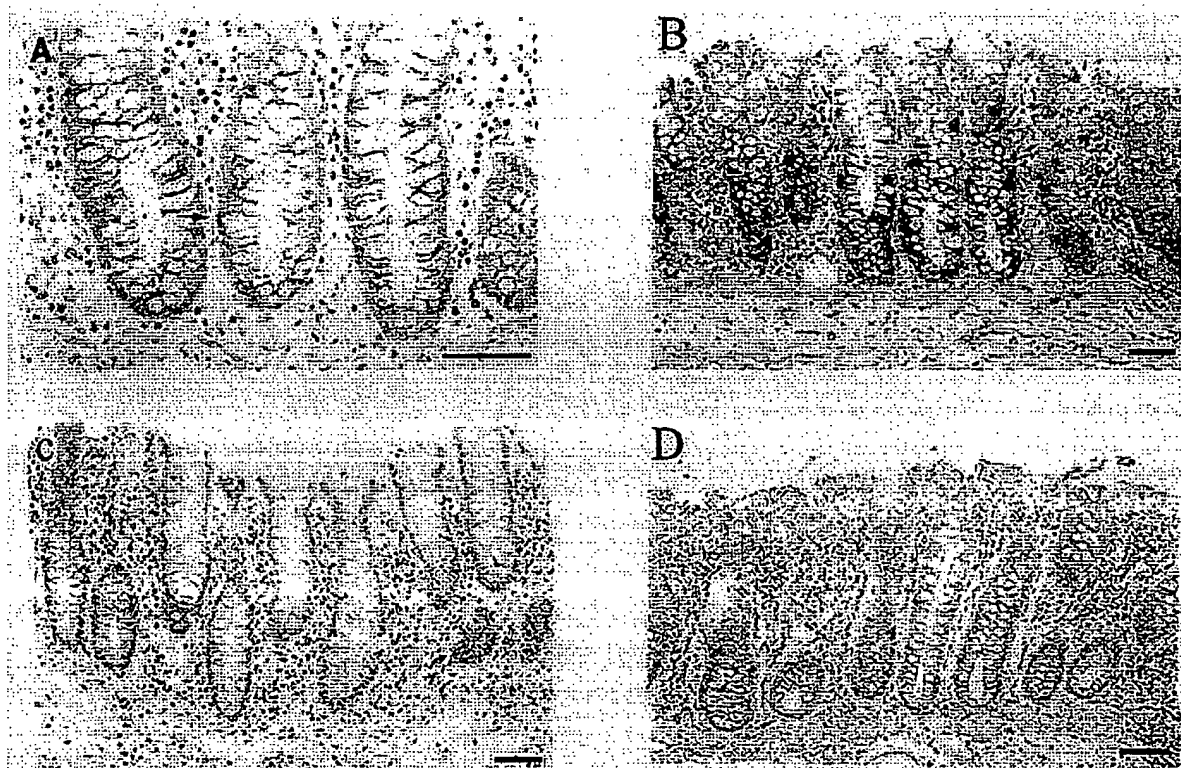


Figure 1
Normal colonic mucosa showing positive nuclear immunostaining for: (a) *cyclin D1*, (b) ISH of *histone H3* mRNA, (c) *Ki-67* and (d) *cyclin A*

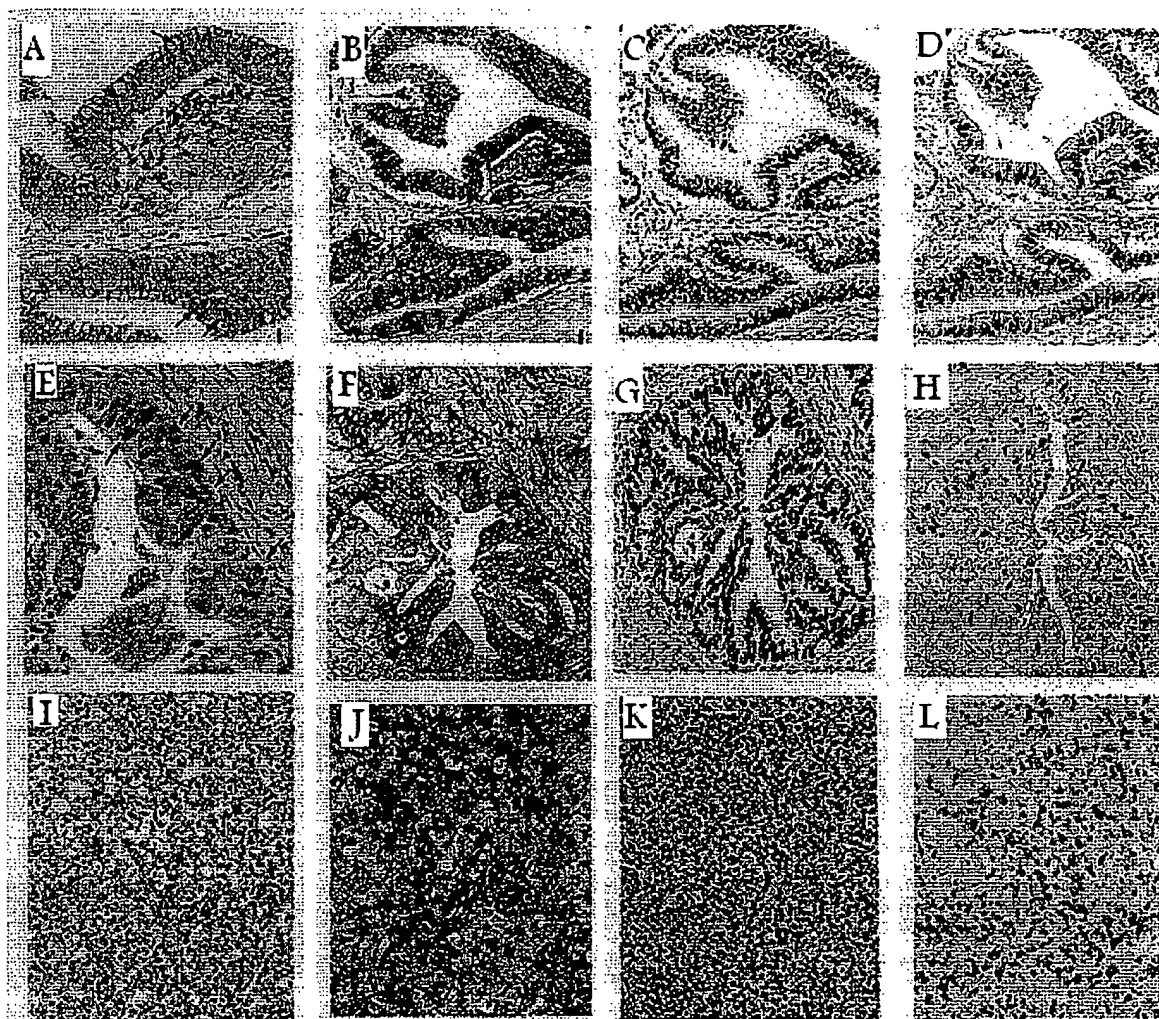
detected in the neoplastic cells of well and moderately-differentiated adenocarcinomas whereas a diffuse homogeneous staining pattern was detected in poorly-differentiated carcinomas. The SI ranged from 10–40.2 (mean: 24.6 ± 6.5).

Immunostaining for *cyclin D1* was predominantly nuclear but cytoplasmic staining was detected in some cases. However, unless a nuclear staining was also detected, cases with cytoplasmic staining were considered negative. Normal colorectal mucosal samples were almost negative for *cyclin D1* whereas 41 out of the 60 (68.3%) CRC cases were positive. Marked heterogeneity was observed in well- and moderately-differentiated adenocarcinomas even within the same tumor. Poorly-differentiated carcinomas revealed a diffuse staining pattern with more darkly-stained nuclei. The SI ranged from 0.5–28.6 (mean: 9.3 ± 4.2).

Positive nuclear staining for *cyclin A* was detected in 80% (48/60) of CRC cases and in all non-neoplastic control samples. Positively-stained nuclei were confined to the lower half of the crypts in normal colonic mucosa and diffusely-dispersed in carcinomas. The SI ranged from 3.3–30.2 (mean: 15.1 ± 6.6).

Histone H3 mRNA was intensely expressed in the cytoplasm of all examined samples either neoplastic or non-neoplastic. The distribution of *histone H3* mRNA was similar to that of *cyclin A* and *Ki-67* however, the proportion of *histone H3* mRNA positive cells was less than that of *Ki-67*. The SI ranged from 1.8–24.2 (mean: 12.4 ± 5.3).

The *PRAD-1* probe detected 3 *EcoRI* fragments of 4.0, 2.2 and 2.0 and 1 *BglII* fragment of 15 Kb. *PRAD-1/cyclin D1* gene amplification was detected in 22/50 (44%) cases analyzed. The degree of amplification was heterogeneous

**Figure 2**

A case of well differentiated adenocarcinoma with positive immunostaining for: (a) *cyclin D1*, (b) *histone H3* mRNA, (c) *Ki-67*, and (d) *cyclin A*. Another case of moderately differentiated adenocarcinoma with positive immunostaining for: (e) *cyclin D1*, (f) *histone H3* mRNA, (g) *Ki-67*, and (h) *cyclin A*. A case of poorly differentiated adenocarcinoma with diffuse staining for: (i) *cyclin D1*, (j) ISH of *histone H3* mRNA, (k) *Ki-67* and (l) *cyclin A*.

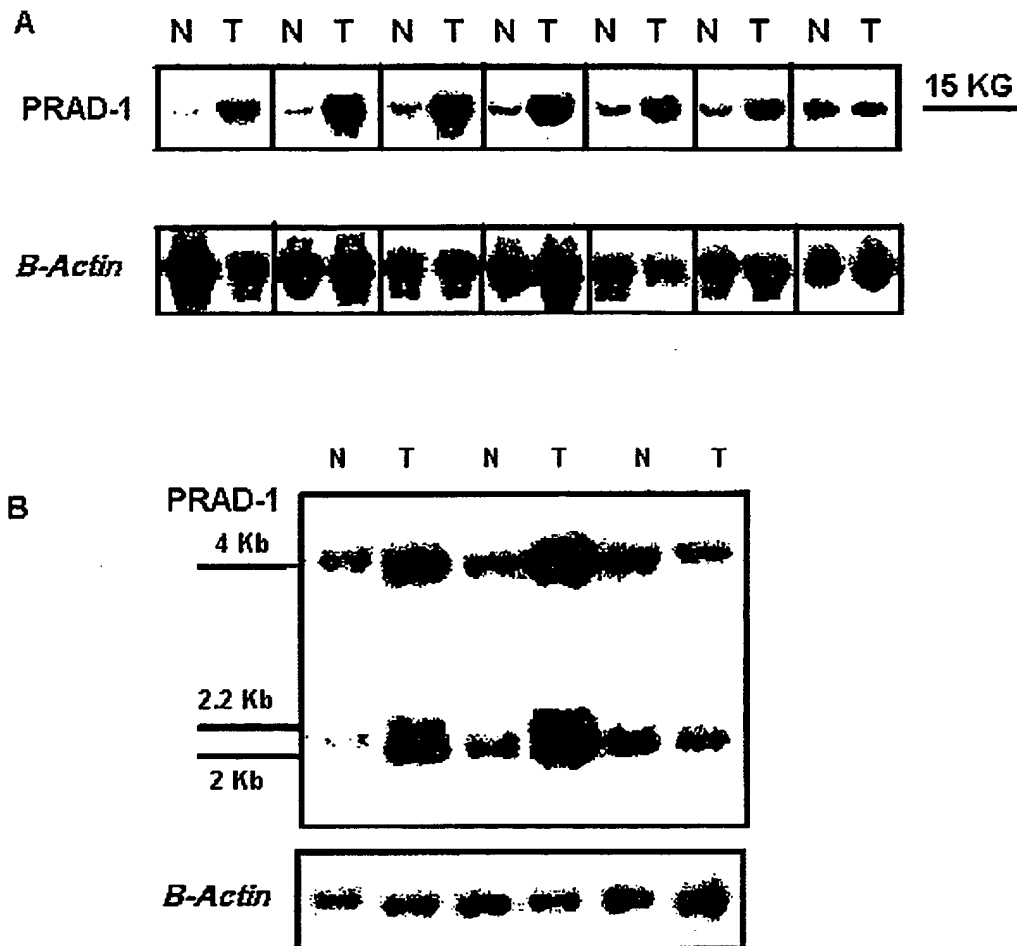
with 2–10 fold increase when compared to normal mucosal samples (Figure 3). Amplification was confirmed by other restriction enzymes.

Correlations

There was a significant correlation between *cyclin D1* gene amplification and protein overexpression. Out of the 22

cases that showed amplification 14 showed protein overexpression (concordance = 63.6%).

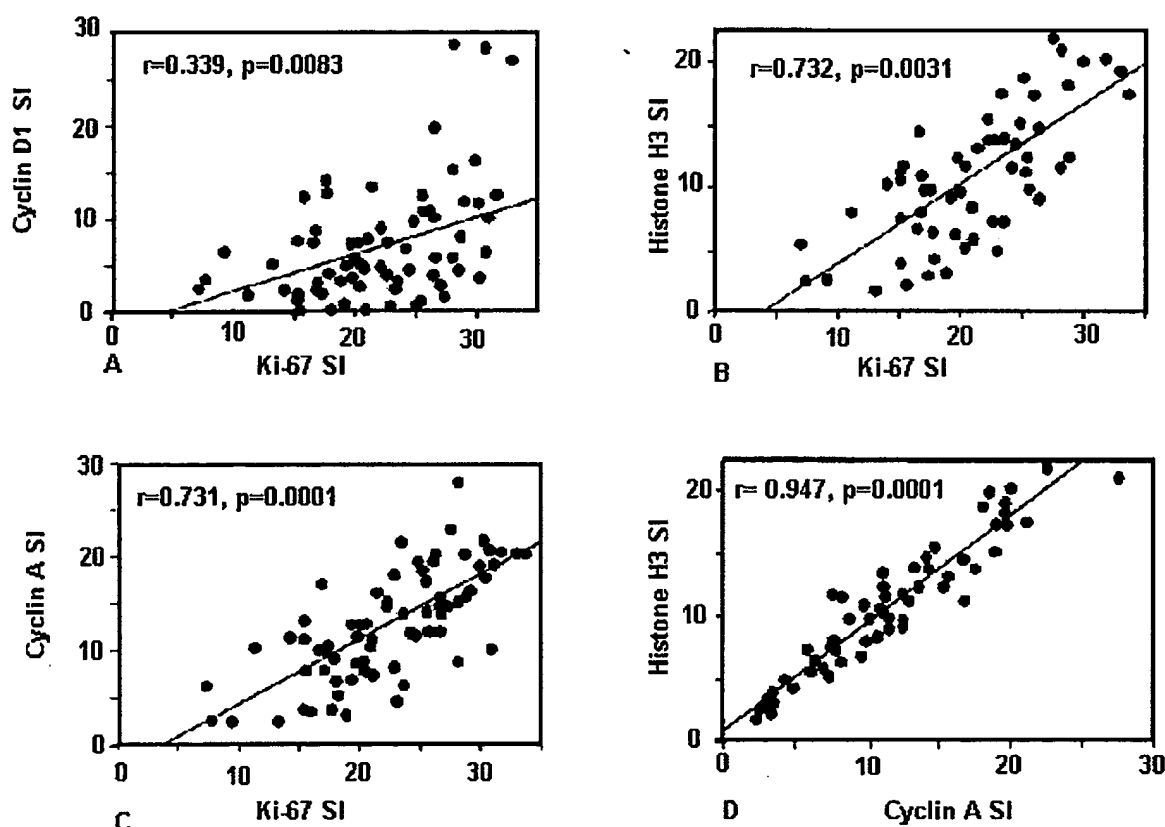
Linear regression analysis of SIs revealed a significant correlation between *Ki-67* and *cyclin D1*, *cyclin A*, *histone H3* as well as between the SIs of *cyclin A* and *histone H3* ($p = 0.008$, 0.0001 , and 0.0001 respectively) (Figure 4). There was a significant relationship between the SI of both *Ki-67*

**Figure 3**

A: Southern blot analysis of normal mucosa (N) and their seven corresponding cases of colonic adenocarcinomas (T1-T7), cases No. 1, 2, 4, and 5 are poorly differentiated whereas cases No. 3, 6, and 7 are moderately differentiated. Genomic DNA was digested with *Bgl*II, fractionated by electrophoresis in agarose gel, transferred onto membranes and hybridized with *PRAD1* and β -actin. Tumors number 1-6 (Lanes 1-6) show different degrees of *PRAD1/cyclin D1* amplification, tumor number 7 (lane 7) was not amplified. **B:** Southern blot analysis of 3 cases of adenocarcinomas (T) and matched normal colonic mucosa (N). Genomic DNA was digested with *Eco*RI, fractionated by electrophoresis in agarose gel, transferred onto membranes and hybridized with *PRAD1* and β -actin probes for loading control. The identification of the 3 tumors is the same as in Fig. 3A with amplification of *PRAD1/cyclin D1* in tumors number 4, 5 (Lanes 1, 2) but not 7 (Lane 3).

and *cyclin A* and the degree of differentiation of tumors as well as the size of the tumor ($p < 0.001$ and $p < 0.01$ respectively). In addition, SI of *Ki-67* and *histone H3* were higher in patients <50 years than in those ≥ 50 years ($p < 0.05$) (table 1).

In addition table 2 shows a significant relationship between high *cyclin D1* SI and large, poorly-differentiated tumors, carcinomas with positive lymph node metastasis and deeply-invasive carcinomas ($p < 0.05$, $p < 0.001$, $p < 0.05$ and $p < 0.05$ respectively). Whereas *cyclin D1* gene amplification was significantly associated with an advanced disease stage since amplification was detected in

**Figure 4**

Correlation between the staining intensity of (a) Ki-67 vs. cyclin D1, (b) Ki-67 vs. histone H3, (c) Ki-67 vs. cyclin A and (d) cyclin A vs. histone H3 mRNA expression.

10/15 (66.7%) of stage IV tumors compared to 12/45 (26.7%) of stage I-III tumors ($p = 0.002$). Similarly, DNA amplification was detected in 60.5% (26/43) of the carcinomas with extensive local invasion (beyond sm) but only in 23.5% (4/17) of the carcinomas with limited invasion (m, sm) ($p = 0.001$). A significant correlation was also present between *cyclin D1* gene amplification and the presence of lymph node metastasis ($p = 0.008$) as well as between the SI of *histone H3*, the size of the tumor and the patient's age ($p < 0.05$, $p < 0.001$ respectively). The SI was higher in tumors >5 cm in diameter and in patients <50 years.

Survival analysis

The mean follow-up period for all patients was 30 months (range: 1–66 months). Eighteen of 60 patients had already died by the time the study was completed. We

defined the cutoff level for overexpression of each cell cycle marker at the point that showed the maximum difference of survival rate between the 2 groups separated by that point. Cox regression analysis revealed that *cyclin A* overexpression (our definition: $SI \geq 10.5$), *cyclin D1* overexpression (our definition: $SI \geq 6.1$), poorly differentiated histology, lymph node metastasis, TNM stage, tumor size and depth of invasion were all significant prognostic variables for survival (Table 3). The Kaplan-Meier survival curves for the subgroups of patients who are subdivided according to each marker's status are shown in Figure 5. Patient with tumors that showed Ki-67 overexpression (our definition: $SI \geq 11.5$) and *histone H3* overexpression (our definition: $SI \geq 8.2$) tended to have poor prognosis but this did not reach a statistically significant level, however the overall survival was significantly lower in patient with *cyclin A* and *cyclin D1* overexpression. Cox multivari-

Table 2: The relation between cyclin D1 overexpression vs cyclin D1 amplification and clinicopathological prognostic markers.

Variables	No. of cases	Cyclin D1 overexpression	Cyclin D1 Amplification
Tumor size (cm)			
<5.0	33	5.3 ± 3.8*	13/33
≥5.0	27	22.8 ± 7.2 p <0.05	9/27 p <0.236
Histology			
GI	15	6.6 ± 4.0	7/15
GII	21	8.9 ± 3.6	8/21
GIII	24	22.0 ± 8.1 p <0.001	7/24 p <0.075
Lymph node			
Negative	33	5.4 ± 5.3*	6/33 (18.2%)
Positive	27	20.6 ± 6.9 p <0.05	16/27 (59.3%) p <0.008
Depth of invasion			
m, sm	17	3.1 ± 3.1*	4/17 (23.5%)
beyond sm	43	12.4 ± 6.5 p <0.05	26/43 (60.5%) p <0.001
Stage			
early	45	5.5 ± 10.1	12/45 (26.7%)
late	15	11.3 ± 9.6 P = 0.175	10/15 (66.7%) p <0.002

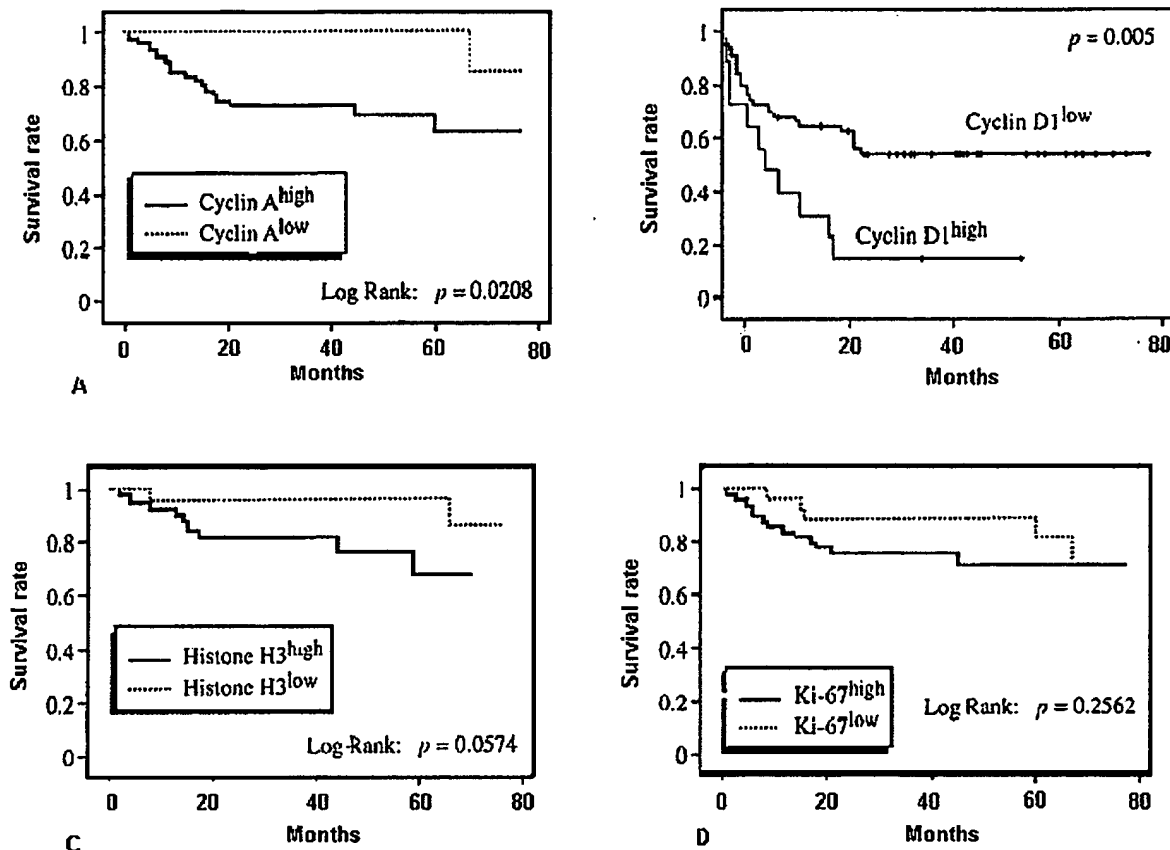
Table 3: Univariate analysis of the relationship between survival and the tested markers

Predictive Variables	Median Survival	HR	CI	P
Ki-67				
<11.5	36			
≥11.5	32	1.826	0.636 – 5.243	0.26
Cyclin D1				
<6.1	35			
≥6.1	18	7.246	1.007 – 45.150	0.03*
Histone H3				
<8.2	35			
≥8.2	29	4.639	0.854 – 25.196	0.07
Cyclin A				
<10.5	35			
≥10.5	15	7.820	1.017 – 60.122	0.02*
Histological grade				
Low	38			
High	10	7.331	2.696 – 19.940	0.0001*
Lymph node				
Negative	38			
Positive	15	6.826	1.973 – 23.621	0.002*
Stage				
I, II, III	38			
IV	12	6.378	1.842 – 22.083	0.001*
Tumor size (cm)				
<5.0	35			
≥5.0	13	4.835	1.386 – 16.868	0.01*
Depth of invasion				
T1, T2	36			
T3, T4	20	7.759	1.024 – 58.789	0.04*
Age (years)				
<50	38			
≥50	28	2.802	0.988 – 7.943	0.0526
Sex				
Male	38			
Female	36	0.696	0.0274 – 1.766	0.4449

* p. value < 0.05 (significant)

HR: Hazard Ratio

CI: 95% confidence Interval

**Figure 5**

Kaplan-Meier survival curves for colorectal carcinoma. Overall survival is significantly lower in patients with (a) *cyclin A* and (b) *cyclin D1* overexpression. Patients with high SI for histone H3 mRNA have poorer prognosis but this was not statistically significant (c). No significant difference was present between patients with high Ki-67 SI and those with low Ki-67 SI (d).

ate regression analysis revealed that lymph node metastasis, *cyclin A* and *cyclin D1* overexpression were independent negative prognostic factors after adjustment for the depth of tumor invasion, age and sex of the patient (Table 4).

Discussion

The proliferative activity of CRC cells has been investigated in several studies either by immunohistochemical determination of cell proliferation index using antibodies to some types of *cyclins* or by flowcytometric determination of the SPF of the cell cycle [8]. Although Leach et al. [17] did not find *cyclin D1* gene amplification in a panel of 47 CRC cell lines; its protein was overexpressed in about 30% of CRC cases that were included in the studies

of Bartakova et al. [6] and Arber et al. [18]. In the former study [6] *cyclin D1* was aberrantly accumulated in a significant subset of human CRC cases and the cell lines derived from these cases were dependent on *cyclin* in their cell cycle progression. In the second study [18], overexpression of *cyclin D1* was detected in 30% of adenomatous polyps indicating that overexpression is a relatively early event in colon carcinogenesis which is possibly responsible for the pathological changes in the mucosa preceding neoplastic transformation. More recently, Holland et al. [19], Pasz-Walczak et al. [20] and Utsunomiya et al. [21] reported up-regulation of *cyclin D1* in 58.7%, 100% and 43% of their studied cases respectively.

Table 4: Multivariate analysis of the relationship between survival and the tested markers

Predictive Variables	HR	CI	P
Cyclin D1	10.864	1.055 – 86.250	0.03*
(baseline < 6.1)	-	-	-
Cyclin A	13.886	1.012 – 190.579	0.0490*
(baseline < 10.5)	-	-	-
Positive Lymph node metastasis	3.921	1.057 – 14.472	0.0410*
Stage IV	3.411	1.048 – 12.083	0.03*
Depth of Invasion			
T3, T4	5.408	0.449 – 65.080	0.1836
Age (years)			
≥50	1.996	0.678 – 5.878	0.2310
Sex	0.910	0.315 – 2.358	0.8453

p. value < 0.05 (significant)
 HR: Hazard Ratio
 CI: 95% confidence Interval

In the present study, up-regulation of *cyclin D1* was detected in 68.3% of the cases. The SI was significantly higher in carcinomas than in normal colorectal mucosa and in poorly-differentiated adenocarcinomas it was approximately twice that of other histological types. Amplification and/or overexpression of *cyclin D1* significantly correlated with deeply invasive tumors and positive lymph node metastasis. Our results in this regards are consistent with previous studies [8,22]. In 2001, Holland et al. [19] demonstrated that deregulation of *cyclin D1* and *p21^{waf}* proteins are important in colorectal tumorigenesis and have implications for patient prognosis. Similarly McKay et al. [23] found that *cyclin D1* was the only protein in their panel (*cyclin D1*, *p53*, *p16*, *Rb-1*, *PCNA* and *p27*) that correlated with improved outcome in CRC patients. However, few studies failed to detect any correlation between *cyclin D1* overexpression and the clinicopathological factors in CRC [6,18]. This controversy in results could partially be explained by the difference in the sampling of studied cases. The present study included 24 cases of poorly differentiated adenocarcinoma, which is not common in other studies of CRC in western countries. This was possible because the majority of CRC cases diagnosed in Egypt are of high histological grade [3]. The correlation between *cyclin D1* overexpression and the high histological grade was also reported in other tumor types including non-small cell lung carcinomas [24] and squamous cell carcinomas of the larynx [16]. Another possible explanation for the observed controversy in the results of different studies is the detection method used.

In the present work, overexpression of *cyclin D1* was more common than gene amplification of the *PRAD-1/cyclin D1*

gene with a 63.6% concordance. This was similarly reported by Bartakova et al. [6] who mentioned that there is a subset of CRC cases in which *cyclin D1* is overexpressed without *PRAD-1/cyclin D1* gene amplification. Consistent with this hypothesis are reports of elevated *cyclin D1* mRNA levels and immunohistochemically detectable accumulation of the protein in over one third of breast cancer cases at a frequency significantly higher than that deduced from DNA amplification studies [9,25]. These data imply that mechanisms other than gene amplification can also lead to deregulation and accumulation of *cyclin D1* in solid tumors.

So far, several studies were done to reveal the prognostic significance of *cyclin D1* overexpression in various carcinomas, including CRC [22]. However, these studies yielded conflicting results which could be attributed to organ heterogeneity. In our study, patients with tumors that exhibited *cyclin D1* overexpression tended to have poor prognosis.

It was reported that, patients with *cyclin A* positive carcinomas had significantly shorter median survival times. Handa et al. [8] were able to detect *cyclin A* overexpression in 77% of their CRC cases. They also demonstrated that, *cyclin A* could be used as a prognostic factor of CRC. More recently, Habermann et al. [26] studied cases of ulcerative colitis with and without an associated adenocarcinoma for the presence of *cyclin A* overexpression. They found that, *cyclin A* overexpression was higher in cases of ulcerative colitis with adenocarcinomas than in those without adenocarcinomas. Consequently, they concluded that, *cyclin A* could be used for monitoring ulcerative colitis patients and for the early detection of an emerging carcinoma in this high risk group of patients.

In our study, *cyclin A* was detected in 80% of the patients and Cox regression analysis showed that it could be used as a prognostic marker in CRC in addition to *cyclin D1*.

It would have been useful if we assessed the expression level of *cyclin A* by another technique (DNA amplification). This would have added more information regarding the gene status on one hand and confirmed the results of IHC on the other hand. Unfortunately, this was not possible because in most of the cases included in the present work, the extracted DNA was not sufficient to study *cyclin amplification* after the assessment of *cyclin D1*.

In 1996, Nagao et al. [11] reported that *histone H3* labeling index significantly correlated with ki-67 immunostaining and was high in poorly differentiated human hepatocellular carcinoma. This was similarly reported in the present work since we found a significant correlation between the SI of *histone H3* and Ki-67. However, no

statistically significant correlation was found between *histone H3* SI and any of the studied clinicopathological factors.

Although *Ki-67* immunostaining reflects the proliferative activity of CRC, it has not been recognized as a significant prognostic factor in this type of tumors [27,28]. However, Suzuki et al. [29] found a significant correlation between *Ki-67* labeling index and local invasion of CRC. In the present study there was a significant relationship between the SI of *Ki-67*, tumor size and grade. However, Kaplan-Meier survival curves showed no significant difference in survival rates between patients with- and without overexpression of *Ki-67*.

Conclusions

Our results demonstrate that *cyclin D1*, *cyclin A*, *histone H3* and *Ki-67* are overexpressed in a subset of CRC, however only *cyclin D1* and *cyclin A* overexpression correlates with poor differentiation and tumor progression. This indicates the superiority of *cyclin A* and *cyclin D1* as indicators of poor prognosis compared to *Ki-67* and *histone H3* mRNA in CRC. *Cyclin A* and *D1* could therefore be considered significant, independent prognostic factors in CRC patients. These findings are especially important in stage II patients since 25–30% of those patients have poor prognosis in spite of being node-negative. However, the standard clinicopathologic prognostic factors can not identify this subset accurately and therefore; there is a great demand for more accurate, individually-based, biological prognostic parameters that help in detecting this high risk group of patients who can benefit from an adjuvant therapy. If the findings of the present study are confirmed in a larger study, evaluation of *cyclin A* and *D1* may be applicable to clinical management of CRC, allowing the identification of patients with poor prognosis.

Competing interests

The author(s) declare that they have no competing interests.

List of abbreviations

CRC – Colorectal cancer

OS – overall survival

SI – staining index

SPF – S phase fraction

ISH – in situ hybridization

m – muscularis mucosa

sm – invasion of the sub mucosa

Authors' contributions

BA and ZA-R carried out the molecular genetic studies, designed, coordinated the study and drafted the manuscript. BA and El-HS carried out all the histopathological and immunohistochemical studies. El-SA participated in molecular genetic studies and drafted the manuscript. MM coordinated the study. El-SM carried out all the patient clinical data. All authors read and approved the final manuscript

References

1. Jiang GL, Huang S: Adenovirus expressing RIZ1 in tumor suppressor gene therapy of microsatellite unstable colorectal cancers. *Cancer Res* 2001, 61:1796-1798.
2. Soliman AS, Bondy ML, Levin B, Hamza MR, Ismail K, Ismail S, Hamam HM, El-Hattab O, Kamal SM, Soliman AG, Dorgham LA, McPherson RS, Beasley RP: Colorectal cancer in Egyptian patients under 40 years of age. *Int J Cancer* 1997, 71:26-30.
3. Soliman AS, Bondy ML, Guan Y, El-Badawy S, Mokhtar N, Bayomi S, Raouf AA, Ismail S, McPherson RS, Abdel-Hakim TF, Beasley PR, Levin B, Wei Q: Reduced expression of mismatch repair genes in colorectal cancer patients in Egypt. *Int J Oncol* 1998, 12:1315-1319.
4. Cordon-Cardo C: Mutations of cell cycle regulators. Biological and clinical implications for human neoplasia. *Am J Pathol* 1995, 147:545-560.
5. Hunter T, Pines J: Cyclins and cancer. II. Cyclin D and CDK inhibitors come of age. *Cell* 1994, 79:573-578.
6. Bartkova J, Lukas J, Strauss M, Bartek J: The PRAD-1/cyclin D1 oncogene product accumulates aberrantly in a subset of CRCs. *Int J Cancer* 1994, 58:568-573.
7. Motokura T, Arnold A: Cyclins and oncogenesis. *Biochim Biophys Acta* 1993, 1155:63-78.
8. Handa K, Yamakawa M, Takeda H, Kimura S, Takahashi T: Expression of the cell cycle markers in colorectal carcinoma: Superiority of cyclin A as an indicator of poor prognosis. *Int J Cancer* 1999, 84:225-233.
9. Gillett C, Fand V, Smith R, Fisher C, Bartek J, Dickson C, Barnes D, Peters G: Amplification and overexpression of cyclin D1 in breast cancer detected by immunohistochemical staining. *Cancer Res* 1994, 54:1812-1817.
10. Gown AM, Jiang JJ, Matles H, Skelly M, Goodpaster T, Cass L, Reshatof M, Spaulding D, Coltrera DM: Validation of the S-phase specificity of histone (H3) in situ hybridization in normal and malignant cells. *J Histochem Cytochem* 1996, 44:221-226.
11. Nagao T, Ishida Y, Kondo Y: Determination of S-phase cells by in situ hybridization for histone H3 mRNA in hepatocellular carcinoma: correlation with histological grade and other cell proliferative markers. *Mod Pathol* 1996, 9:99-104.
12. Chou MY, Chang AL, McBride J, Donoff B, Gallagher GT, Wong DT: A rapid method to determine proliferation patterns of normal and malignant tissues by H3 mRNA in situ hybridization. *Am J Pathol* 1990, 136:729-733.
13. Sobin LH, Wittekind C: TNM classification of malignant tumors. 5th edition. John Wiley, New York; 1997.
14. King RJ, Coffey AI, Gilbert J, Lewis K, Nash R, Millis R, Raju S, Taylor RW: Histochemical studies with a monoclonal antibody raised against a partially purified soluble estradiol receptor preparation from human myometrium. *Cancer Res* 1985, 45:5728-5733.
15. Slebos RJ, Boerrigter L, Evers SG, Wisman P, Mooi WJ, Rodenhuis S: A rapid and simple procedure for the routine detection of ras point mutations in formalin-fixed, paraffin-embedded tissues. *Diag Mol Path* 1992, 1:136-141.
16. Jares P, Fernandez P, Campo E, Nadal A, Bosch F, Aiza G, Nayach I, Traserra J, Cardesa A: PRAD-1/cyclin D1 gene amplification correlates with messenger RNA overexpression and tumor progression in human laryngeal carcinomas. *Cancer Res* 1994, 54:4813-4817.
17. Leach FS, Elledge SJ, Sherr CJ, Willson JK, Markowitz S, Kinzler KW, Vogelstein B: Amplification of cyclin genes in colorectal carcinomas. *Cancer Res* 1993, 53:1986-1989.

18. Arber N, Hibshoosh H, Moss SF, Sutter T, Zhang Y, Begg M, Wang S, Weinstein IB, Holt PR: Increased expression of cyclin D1 is an early event in multistage colorectal carcinogenesis. *Gastroenterology* 1996, 110:669-674.
19. Holland TA, Elder J, McCloud JM, Hall C, Deakin M, Fryer AA, Elder JB, Hoban PR: Subcellular localization of cyclin D1 protein in colorectal tumors is associated with p21 (WAF1/CIP1) expression and correlates with patient survival. *Int J Cancer* 2001, 95(5):302-306.
20. Pasz-Walczak G, Kordek R, Falik M: P21(WAF1) expression in colorectal cancer: correlation with p53 and cyclin D1 expression, clinicopathological parameters and prognosis. *Pathol Res Pract* 2001, 197(10):683-689.
21. Utsunomiya T, Doki Y, Takemoto H, Shiozaki H, Yano M, Sekimoto M, Tamura S, Yasuda T, Fujiwara Y, Monden M: Correlation of beta-catenin and cyclin D1 expression in colon cancers. *Oncology* 2001, 61(3):226-233.
22. Maeda K, Chung YS, Kang SM, Ogawa M, Onoda N, Nakata B, Nishiguchi Y, Ikehara T, Okuno M, Sowa M: Overexpression of cyclin D1 and p53 is associated with disease recurrence in colorectal adenocarcinoma. *Int J Cancer* 1997, 74:310-315.
23. McKay JA, Douglas JJ, Ross VG, Curran S, Loane JF, Ahmed FY, Cassidy J, McLeod HL, Murray GI: Analysis of key cell cycle checkpoint proteins in colorectal tumors. *J Pathol* 2002, 196:386-393.
24. Mate JL, Ariza A, Aracil C, Lopez D, Isamat M, Perez-Pitella J, Navas-Palacios JJ: Cyclin D1 overexpression in non-small cell lung carcinoma: correlation with Ki-67 labeling index and poor cytoplasmic differentiation. *J Pathol* 1996, 180:395-399.
25. Buckley MF, Sweeney KJ, Hamilton JA, Sini RL, Manning DL, Nicholson RI, DeFazio A, Watts CK, Musgrove EA, Sutherland RL: Expression and amplification of cyclin genes in human breast cancer. *Oncogene* 1993, 8:2127-2133.
26. Habermann J, Lenander C, Roblick UJ, Kruger S, Ludwig D, Alalya A, Freitag S, Dumbgen L, Bruch HP, Stange E, Salo S, Tryggvason K, Auer G, Schimmelpenninck H: Ulcerative colitis and colorectal carcinoma: DNA profile, laminin-5 gamma 2 chain and cyclin A expression as early markers for risk assessment. *Scand J Gastroenterol* 2001, 36:751-758.
27. Kubota Y, Petras RE, Easley KA, Bauer TW, Tubbe RR, Fazio VW: Ki-67-determined growth fraction versus standard staging and grading parameters in colorectal carcinoma. A multivariate analysis. *Cancer* 1992, 70:2602-2609.
28. Shain AA, Ro JY, Brown RW, Ordonez NG, Cleary KR, El-Naggar AK, Wilson P, Ayala AG: Assessment of Ki-67-derived tumor proliferative activity in colorectal adenocarcinomas. *Mod Pathol* 1994, 7:17-22.
29. Suzuki H, Matsumoto K, Terabe M: Ki-67 antibody labeling index in colorectal carcinoma. *J Clin Gastroenterol* 1992, 15:317-320.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-230X/4/22/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



BioMed Central

Correlation of amplification and overexpression of the *c-myc* oncogene in high-grade breast cancer: FISH, *in situ* hybridisation and immunohistochemical analyses

J Blancato^{*,1,2}, B Singh², A Liu³, DJ Liao⁴ and RB Dickson^{2,5}

¹Institute for Molecular and Human Genetics, 3970 Reservoir Road, NW, Washington DC 20007, USA; ²Lombardi Comprehensive Cancer Center, 3970 Reservoir Road, NW, Washington DC 20007, USA; ³Biometry and Mathematical Statistics Branch, National Institute of Child Health and Human Development, 9000 Rockville Pike, Bethesda, MD 20892-7510, USA; ⁴Department of Radiation Oncology, Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA; ⁵Department of Oncology, Georgetown University Medical Center, 3970 Reservoir Road, NW, Washington DC 20007, USA

In this study, we analysed gene amplification, RNA expression and protein expression of the *c-myc* gene on archival tissue specimens of high-grade human breast cancer, using fluorescent *in situ* hybridisation (FISH), nonradioactive *in situ* hybridisation and immunohistochemistry. The specific question that we addressed was whether expression of *c-Myc* mRNA and protein were correlated with its gene copy amplification, as determined by FISH. Although *c-Myc* is one of the most commonly amplified oncogenes in human breast cancer, few studies have utilised *in situ* approaches to directly analyse the gene copy amplification, RNA transcription and protein expression on human breast tumour tissue sections. We now report that by using the sensitive FISH technique, a high proportion (70%) of high-grade breast carcinoma were amplified for the *c-myc* gene, irrespective of status of the oestrogen receptor. However, the level of amplification was low, ranging between one and four copies of gene gains, and the majority (84%) of the cases with this gene amplification gained only one to two copies. Approximately 92% of the cases were positive for *c-myc* RNA transcription, and essentially all demonstrated *c-myc* protein expression. In fact, a wide range of expression levels were detected. Statistically significant correlations were identified among the gene amplification indices, the RNA expression scores and protein expression scores. *c-myc* gene amplification, as detected by FISH, was significantly associated with expression of its mRNA, as measured by the intensity of *in situ* hybridisation in invasive cells ($P = 0.0067$), and by the percentage of invasive cells positive for mRNA expression ($P = 0.0006$). *c-myc* gene amplification was also correlated with the percentage of tumour cells which expressed high levels of its protein, as detected by immunohistochemistry in invasive cells ($P = 0.0016$). Thus, although multiple mechanisms are known to regulate normal and aberrant expression of *c-myc*, in this study, where *in situ* methodologies were used to evaluate high-grade human breast cancers, gene amplification of *c-myc* appears to play a key role in regulating expression of its mRNA and protein. *British Journal of Cancer* (2004) 90, 1612–1619. doi:10.1038/sj.bjc.6601703 www.bjcancer.com

Published online 30 March 2004

© 2004 Cancer Research UK

Keywords: *c-myc*; breast cancer; gene amplification; gene expression

The *c-myc* oncogene has been shown to be amplified and/or overexpressed in many types of human cancer (Marcu *et al*, 1992; Nass and Dickson, 1997; Nesbit *et al*, 1999; Liao and Dickson, 2000). Numerous experiments *in vivo* have also causally linked aberrant expression of this gene to the development and progression of cancer in different body sites (Marcu *et al*, 1992; Nass and Dickson, 1997; Nesbit *et al*, 1999; Liao and Dickson, 2000). However, several critical issues regarding the significance of *c-myc* in human cancer still remain obscure. First, even for a given type of malignancy, the frequencies of the alterations of *c-myc* at the cytogenetic and expression levels vary greatly from one report to another (Liao and Dickson, 2000). For instance, the frequencies of its amplification, mRNA and protein overexpression in breast cancer vary between 1–94, 22–95 and roughly 50–100%, respectively, among different reports (Liao and Dickson, 2000).

Thus, it is still unclear to what extent this gene is altered at the cytogenetic level and at different expression levels in breast carcinoma.

One controversial issue pertains to the prognostic value of *c-myc* gene alterations in cancer. The central role of *c-Myc* protein in accelerating cell proliferation, documented by many early studies, has led to a general concept for many types of cancer that amplification or overexpression of this gene may be associated with a more aggressive tumour and a poorer patient survival (Berns *et al*, 1992; Marcu *et al*, 1992; Sato *et al*, 1995; Nass and Dickson, 1997; Nesbit *et al*, 1999; Visca *et al*, 1999; Liao and Dickson, 2000). However, many reports have shown an opposite correlation (Sikora *et al*, 1985, 1987; Watson *et al*, 1986; Polaczar *et al*, 1989; Voravud *et al*, 1989; Williams *et al*, 1990; Melhem *et al*, 1992; Pietilainen *et al*, 1995; Diebold *et al*, 1996; Smith and Goh, 1996; Augenlich *et al*, 1997; Bieche *et al*, 1999), while other studies do not support either of these conclusions. For instance, gene amplification or overexpression of *c-Myc* protein has also been shown to associate with a better tumour differentiation or a better

*Correspondence: Dr J Blancato; E-mail: blancatj@georgetown.edu
Received 7 July 2003; revised 17 January 2004; accepted 17 January 2004;
published online 30 March 2004

patient survival for cancer of the testis, ovary, bile ducts, colon and breast (Sikora *et al.*, 1985, 1987; Watson *et al.*, 1986; Polaczar *et al.*, 1989; Voravud *et al.*, 1989; Williams *et al.*, 1990; Melhem *et al.*, 1992; Pietilainen *et al.*, 1995; Diebold *et al.*, 1996; Smith and Goh, 1996; Augenlich *et al.*, 1997; Bieche *et al.*, 1999). This controversy does not appear to be related completely to the cancer type, since both positive (Berns *et al.*, 1992; Visca *et al.*, 1999) and negative (Williams *et al.*, 1990; Melhem *et al.*, 1992; Pietilainen *et al.*, 1995; Smith and Goh, 1996; Augenlich *et al.*, 1997; Bieche *et al.*, 1999) correlations have been reported for colon cancer and breast cancer. More interestingly, *c-Myc* overexpression has been shown to predict a poorer prognosis for cutaneous melanoma, but a favourable outcome for uveal melanoma (Grover *et al.*, 1997; Chana *et al.*, 1998a, b, 1999; Grover *et al.*, 1999). These data indicate different roles of *c-Myc*, even in the same type of tumour, perhaps depending upon different tissue microenvironments.

Another controversial issue concerns the nuclear–cytoplasmic localisation of *c-Myc*. Studies of neoplasms of the colon, testis, ovary and liver have shown that predominantly nuclear localisation of *c-Myc* tends to occur in benign lesions, while cytoplasmic localisation tends to occur in more malignant tumours (Sikora *et al.*, 1985; Sundaresan *et al.*, 1987; Melhem *et al.*, 1992; Sasano *et al.*, 1992; Yuen *et al.*, 2001). Whether these patterns of subcellular localisation of *c-Myc* tend to reflect the malignant status of breast cancer remains an enigma.

A recent study of the impact of DNA amplification on gene expression patterns in breast cancer used mRNA and DNA from 14 breast cancer cell lines. Analysis was conducted with a 13 000 cDNA clone array for gene expression measurement and a Comparative Genomic Hybridisation (CGH) microarray for gene copy number measurements. This study also included known breast cancer genes, such as *c-myc*, *HER2-neu* and *ai1* (Hyman *et al.*, 2002). Interestingly, 44% of the most highly amplified genes were also overexpressed at the mRNA level. Consistent with this pattern, *c-Myc* gene copy number and its expression levels showed a statistically significant ($\alpha = 0.020$) correlation in this microarray study of breast cancer cell lines. Another study, by Pollack and colleagues, used microarray analysis and BAC array CGH of RNA and DNA (respectively) extracted from intermediate grade human breast tissues, and tested for amplification and expression of *c-Myc* (among other genes). This study demonstrated that two out of 37 specimens were both amplified and overexpressed, while others were either amplified or overexpressed, but not both. The authors of this study suggested that contaminating stromal tissue may compress the fluorescence ratios leading to underestimates of gene amplification and overexpression (Pollack *et al.*, 2002).

To more clearly address the importance of gene amplification and expression of *c-Myc* in human breast cancer, we used *in situ* methodologies, which can clearly distinguish stromal and carcinoma components. We studied the amplification and overexpression of the *c-myc* gene with fluorescent *in situ* hybridisation (FISH), non-radioactive *in situ* hybridisation (ISH) and immunohistochemical (IHC) approaches on paraffin-embedded biopsy sections of untreated, high-grade breast cancer. It was observed that 70, 92 and 70% of the cancer cases exhibited *c-myc* gene amplification, its mRNA overexpression and its protein overexpression, respectively. In most of the cases (84%) that showed gene amplification, the *c-myc* gene gained only one to two copies, which is consistent with *c-myc* FISH data from other studies. Unlike some oncogenes, such as *N-myc*, which typically demonstrates gene amplification copy numbers of greater than 10 in neuroblastoma, and *HER-2/neu* (Sartelet *et al.*, 2002), whose copy numbers range up to 14–40 in breast carcinomas (Isola *et al.*, 1999), gene copy numbers of *c-myc* are not as greatly increased. In the study noted earlier, using breast cancer cell line CGH array and cDNA microarray expression analysis, it was demonstrated that the most dramatically increased expression levels were associated with large gene copy number increases, although low-level gains

and losses had a significant influence on gene expression dysregulation (Hyman *et al.*, 2002). Only one study has been published (Pollack *et al.*, 2002) that has begun to determine if these findings are directly relevant to actual human breast tumour tissues, since many of the genetic changes in tissue culture cell lines are more extreme than those displayed in primary tumour material. Furthermore, the relationships among gene amplification, mRNA expression and *c-Myc* protein expression were not explored in prior human breast cancer cell line and tumour tissue studies (Hyman *et al.*, 2002; Pollack *et al.*, 2002).

In our human breast tumour tissue study, a high correlation was found between *c-myc* FISH and ISH, for both percentage of staining ($P < 0.0067$) and intensity positive cells ($P < 0.0006$). In addition, *c-myc* gene copy amplification by FISH was correlated with *c-Myc* protein expression positive cells by IHC ($P < 0.0016$). These results support the idea that *c-Myc* overexpression of both mRNA and protein is related to the copy number of the *c-myc* DNA amplification. We show in this study that amplification and overexpression of *c-Myc* occur with high frequency in high-grade human breast cancer tissues.

MATERIALS AND METHODS

Materials

Formalin-fixed, paraffin-embedded tissue blocks of breast carcinoma and normal breast tissue were obtained from the Histopathology and Tissue Shared Resource at the Lombardi Comprehensive Cancer Center (LCCC), at Georgetown University Medical Center. The criteria for tumour selection were the following: negative progesterone receptor status, metastases to auxiliary lymph nodes and high grade (Elston Score > 7). The oestrogen receptor status of the tumours was known from archived pathology reports. The parameters were chosen from our prior meta-analysis (Deming *et al.*, 2000), as indications of a high likelihood of *c-myc* gene amplification. Normal breast tissue specimens were from reduction mammoplasty. Serial sections (5 μ m) for FISH, ISH and IHC were prepared by the LCCC Histopathology and Tissue Shared Resource.

FISH

A dual-label FISH technique was used (Jenkins *et al.*, 1997). Slides were baked overnight at 60°C to assure adherence of the sample. Tissue sections were deparaffinised with two successive, 10 min xylene washes, and then dehydrated in a graded ethanol series of 70, 80 and 95% at room temperature. Samples were then digested with 4% pepsin (Sigma, St Louis, MO, USA) at 45°C for 10 min. DNA probes used were an alpha satellite probe to chromosome 8, labelled with biotin, and a *c-myc* probe, labelled with digoxigenin (Ventana, Tucson, AZ, USA). Codenaturation was performed at 90°C for 10 min on a hot plate. Hybridisation was at 37°C for 12–16 h. Detection of signals was accomplished with an antiavidin antibody labelled with Texas Red, and an antidigoxigenin antibody conjugated to fluorescein (Ventana, Tucson, AZ, USA). Slides were postwashed in $2 \times$ SSC at 72°C for 5 min and counterstained with DAPI to visualise cell nuclei. Results were viewed and quantified with a Zeiss Axiophot fluorescence microscope, equipped with appropriate filters and an Applied Imaging Cytovision system (Pittsburgh, PA, USA). In this approach, the *c-myc* unique sequence probe was visualised as a green signal and the control probe for the chromosome 8 centromere was red, thus easily being distinguished when scored.

One serial section from each tumour sample was stained with haematoxylin and eosin and first reviewed by a pathologist (BS), to help identify the tumour area of the section. This procedure ensured that the tumour cells, but not the normal cells, were

counted. Nuclei of up to 50 tumour cells were scored from each FISH-stained section, independently by two investigators. Hybridisation signals were averaged, and the amplification index was presented as the number of c-myc signals divided by the number of chromosome 8 centromere signals. A 1.8-fold increase was used as the criterion to judge the presence of c-myc gene amplification.

In situ hybridisation

In situ hybridisation (ISH) was carried out with a nonradioactive method, described previously (Liao et al, 2000a, b). One serial section from each specimen was hybridised overnight at 60°C with riboprobes, that were *in vitro* transcribed from the antisense or sense strand of an approximately 300 bp cDNA of human c-myc (ATCC, Manassas, VA, USA), labelled with digoxigenin-conjugated UTP. The sections were then incubated with an antibody against digoxigenin, followed by incubation with a second antibody conjugated to alkaline phosphatase. The signal was visualised by colour development with 5-bromo-4-chloro-3-indolyl phosphate and nitroblue tetrazolium. All reagents were purchased from Boehringer Mannheim, Indianapolis, IA. To control the signal specificity, two serial sections were mounted on the same slide for hybridisation with the antisense and sense probes, respectively. ISH was given an intensity and percentage scores, based on intensity of positive staining and number of cells staining, respectively. Intensity scores were assigned 0, 1, 2 and 3, and percentage scores were assigned as 1-1-25, 2-26-50, 3-51-75 and 4-76-100%.

Immunohistochemistry

Immunohistochemical staining (IHC) was performed using an avidin-biotin complex (ABC) method described previously (Liao et al, 1998). One serial section of each specimen was deparaffinised and blocked with 3% peroxide. Antigens were retrieved by heating slides in a microwave oven in 50 mM citrate buffer, pH 6.4, at boiling temperature, for 12 min. After blocking with 6% normal goat serum, the section was incubated with a mouse monoclonal antibody to human c-Myc (9E10, Sigma Chemical Company, St Louis, MO, USA) at 1:100 dilution for 2 h, followed by 1 h incubation with a second antibody conjugated with biotin (Vector Laboratories Inc., Burlingame, CA, USA). The section was then incubated with peroxidase-conjugated avidin (Dako, Corporation, Carpinteria, CA, USA) for 30 min, followed by colour development with diaminobenzidine and peroxide. All procedures were carried out at room temperature. To control the signal specificity, serial sections from 10 tumour samples were also stained using an alternate c-Myc antibody (C19 from Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA) at 1:60 dilution. This antibody resulted in focally positive staining in the tumour, but the staining intensity was weaker. To control the signal specificity, serial sections were made from five selected positive cases which were subjected to the same staining procedure, with a normal mouse IgG to replace the c-Myc antibody. This control staining did not give rise to a signal, demonstrating the specificity of the c-Myc antibody signal. IHC staining was given an intensity and percentage score based upon the intensity of positive staining and number of cells staining. Intensity scores were assigned 0, 1, 2 and 3 and percentage scores were assigned as 1-1-25, 2-26-50, 3-51-75 and 4-76-100%. Determinations were made of cellular localisation of c-Myc antibody staining to cytoplasm and/or nucleus in normal and invasive cells within each breast tumour specimen.

Statistical analyses

For each analysis of gene copy amplification (FISH), mRNA expression (ISH) and protein expression (IHC), all cases were first grouped as positive or negative to calculate the percentages of

positive cases and negative cases, as described (Zar, 1974). Fisher's exact test was used to compare percentages, and two-sample *t*-test or Wilcoxon rank test was used to compare average scores. Both ISH and IHC were given intensity and percentage scores, based on intensity of positive staining and number of cells staining, respectively. As noted earlier, intensity scores were assigned 0, 1, 2 and 3 and percentage scores were assigned as 1-1-25, 2-26-50, 3-51-75 and 4-76-100%. A score of >2 for either intensity of staining or percentage of cells positive by ISH was assigned as high. For IHC, an intensity score of >1 was assigned as high and a percentage score of >3 was categorised as high. Each amplification index was paired with its corresponding mRNA expression score to calculate the coefficient *r*. The same method was used to estimate the association of the amplification indices with the c-Myc protein expression levels, and the association of the mRNA expression levels with the protein expression levels. A *P*-value of 0.05 or less was used to determine the statistical significance in all analyses. In all, 54 pairs of normal vs invasive tissues were analysed using McNemars χ^2 test to determine if there was a difference in cellular localisation of c-Myc antibody signal to nuclear or cytoplasmic compartments.

RESULTS

FISH analysis of gene amplification

Amplification of the c-myc gene was measured by a FISH test in 46 cases of breast cancer; Figure 1 demonstrates cells with no amplification (one copy of c-myc/one copy of chromosome 8 centromere, and a moderate amplification a 3/1 ratio). Amplification was calculated by the number of c-myc signals divided by the number of chromosome 8 alpha satellite signals. A 1.8-fold increase cut-off was used to judge gene amplification. As shown in Table 1, 32 out of 46 (70%) cases were gene amplified for c-myc, whereas only 30% (14/46) of the cases showed amplification indices lower than the cut-off value. The amplification indices for most (84%, or 27/32) cases with gene amplification, ranged between 1.8- and three-fold, indicating that the locus gained up to two copies of c-myc in the majority of the cases. The percentage of cases with gene gains of three copies or higher was 11% (five out of 46) of total cases analysed, or near 16% (five out of 32) of the cases with gene amplification, including one case (2% of total cases or 3% of the cases with gene amplification) with the highest index of 5 (a gain of four copies).

In all, 28 of the breast carcinomas in this study were ER negative, and 14 were ER positive. The average c-myc gene amplification score was 1.896 (s.e. = 0.196) for ER positive and 2.201 (s.e. = 0.157) for ER negative. Although ER-negative tumours had a slightly higher average c-myc score, the difference was not statistically significant (two-sided *P* = 0.252 from two-sample *t*-test and 0.251 from Wilcoxon rank test), consistent with the results of our prior meta-analysis of the literature (Deming et al, 2000).

In situ hybridisation analysis of c-myc mRNA expression

A total of 51 breast cancer samples were studied for c-Myc mRNA expression, with non radioactive *in situ* hybridisation (ISH). ISH results were assigned intensity and percentage scores based upon signal intensity of positive staining and number of cells staining within the sample, respectively. As shown in Table 2, 86% (44 out of 51) tumours were scored as high in intensity, and 92% (47 out of 51) had more than 51% positive cells, also considered as highly increased c-Myc expression. mRNA expression was heterogeneous in the breast tumour tissue, and no morphologic subtype was predominant in the high or low categories. One case showed no c-Myc ISH staining. In 79% (38/48) of cases, epithelia in normal mammary glands adjacent to the tumour also showed a high

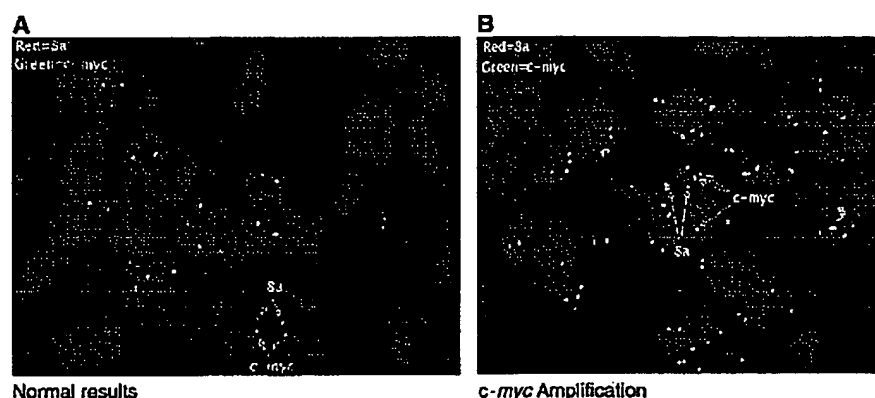


Figure 1 FISH analysis of *c-myc* amplification in tumour cells from breast tumour tissue sections. FISH probe for human *c-myc* unique-sequence is seen as green, while the normal control signal, a centromeric probe signal for chromosome 8 is shown in red. The nuclei of tumour cells were visualised by DAPI counter-staining. (A) 1:1 copy ratio of *c-myc* to chromosome 8 (*c-myc*/8 centromere), indicating no amplification of *c-myc* in tumour cells. (B) 1:3 copy ratio of *c-myc* to chromosome 8 (*c-myc*/8 centromere), a moderate amplification of the *c-myc* gene.

Table 1 *c-myc* gene copy amplification analysis by FISH in poor prognosis human breast tumour samples

Amplification Index (# <i>c-myc</i> signals/# control signals)+	Percentage of samples with FISH ratios in each category	
1.0–1.7	30%	14 out of 46
1.8–1.99	20%	Nine out of 46
2.0–2.9	39%	18 out of 46
>3.0	11%	Five out of 46

Analysis was conducted on 46 individual paraffin-embedded tissue samples with negative progesterone receptor status, positive lymph node involvement and high tumour grade. +Normal control ratio is 1.

intensity of staining. In three cases, no staining was seen in the normal terminal duct lobular units. Figure 2 shows representative fields of high, medium and low *c-myc* mRNA expression levels in invasive ductal carcinoma samples.

Association of FISH and ISH

c-Myc scores were dichotomised as binary variables (high or low), and a score of 2 or higher was categorised as high on ISH. A score higher than median was categorised as high from FISH studies. These dichotomised scores are depicted in Table 3. A Fisher's exact test was performed for comparing binary responses to see if there was any association between FISH and ISH. It was found that the FISH score was significantly associated with percentage of staining in the invasive cells ($P=0.0067$, two-sided McNemar's test) and also with the intensity score on ISH ($P=0.0006$, two-sided).

Immunohistochemical staining of *c-Myc* proteins

In total, 51 breast carcinomas, which were subjected to FISH analysis, and all of which also had been analysed for *c-myc* mRNA by *in situ* hybridisation, were also analysed for the expression of *c-Myc* protein, using immunohistochemical staining with the 9E10 antibody. IHC results were assigned an intensity and percentage score based on intensity of positive staining and number of cells staining, respectively. Intensity scores were assigned 0, 1, 2 and 3 and percentage scores were assigned as 0, 1–25, 26–50, 51–75 and 76–100.

For IHC, an intensity score of >1 was assigned as high and a percentage score of >3 was categorised as high. Figure 2 shows examples of high, medium and low levels of *c-myc* antibody staining in invasive ductal carcinoma samples. In 34 cases, normal tissue was seen; 30 of these showed cytoplasmic staining and 22 had nuclear staining in terminal ductal lobular units. In all, 12 cases showed 1+, 14 cases 2+ and four cases 3+ cytoplasmic staining. *In situ* hybridisation revealed positive staining in 46 out of 49 cases with normal tissue. Seven cases showed 1+, 13 cases showed 2+ and 26 cases showed 3+ staining by ISH. Both immunohistochemistry and *in situ* hybridisation showed diffuse positivity in adipocytes.

Table 4 shows the staining pattern for the cohort. In all, 70% (36 out of 51) of cases showed high intensity of staining for *c-Myc* protein, while 85% (29 out of 34) of cases with detectable staining had more than 76% positive cells, also considered as high expression. To verify the staining specificity, serial sections from 10 tumour specimens that were positive for 9E10 antibody were also stained using the C19 rabbit polyclonal anti-*c-Myc* antibody. Results revealed a staining pattern similar to 9E10. However, the staining intensity with C19 was weaker than 9E10. The specificity of these two antibodies was verified by Western blots in previous studies (Persons *et al*, 1997; Liao *et al*, 2000b). Figure 2 shows results of *c-Myc in situ* hybridisation and immunohistochemistry studies on samples considered to demonstrate low, moderate and high levels of *c-Myc* expression. Analysis of *c-Myc* protein localisation results in the nucleus or cytoplasmic compartments of normal and invasive cells within the tumours revealed that nuclear staining was positive in 41% of normal cells, compared to 22% of invasive cells (statistical significance at $P=0.01$ by McNemar's two-sided χ^2 test). The increase in relative cytoplasmic localisation of *c-Myc* protein, comparing normal (53.7%), to invasive cells (61.1%) was not significantly different. Thus, the data are consistent with partial exclusion of *c-Myc* from the nuclei of invasive breast cancer cells.

The FISH score was significantly associated with the percentage positivity of invasive cells, as seen on IHC studies of *c-Myc*. However, 40% of tumours displayed a low index of *c-myc* gene amplification, but still expressed high levels of *c-Myc* protein (Table 6), indicating the possibility of other mechanisms of over expression unrelated to gene amplification in at least some tumours. The FISH score was not significantly associated with the intensity of IHC staining in the invasive cells (not shown), in contrast to the IHC percentage positivity score.

Table 2 *c-myc* mRNA *in situ* hybridisation (ISH) results

Staining intensity	0	1	2	3	Percent positivity	1	2	3	4
Number of tumour samples in each category N = 51	1	6	25	19	Number of tumour samples in each level category N = 51	1	3	5	42

In all, 51 human high-grade breast carcinomas were analysed to determine the relationships between *c-myc* mRNA expression and *c-myc* gene *in situ* hybridisation results. Data are shown in two ways in the above table. First, overall staining intensity of *c-myc*-positive cells was scored as 0, 1, 2, 3 (low to high), and the number of tumour samples at each level of staining indicated on the line below. Next, the percentage of tumour cells staining was scored as 0, 1, 2, 3, 4 (low to high %, as discussed in Materials and Methods). The number of tumours at each level of percent cell positivity for *c-myc* is then indicated on the line below.

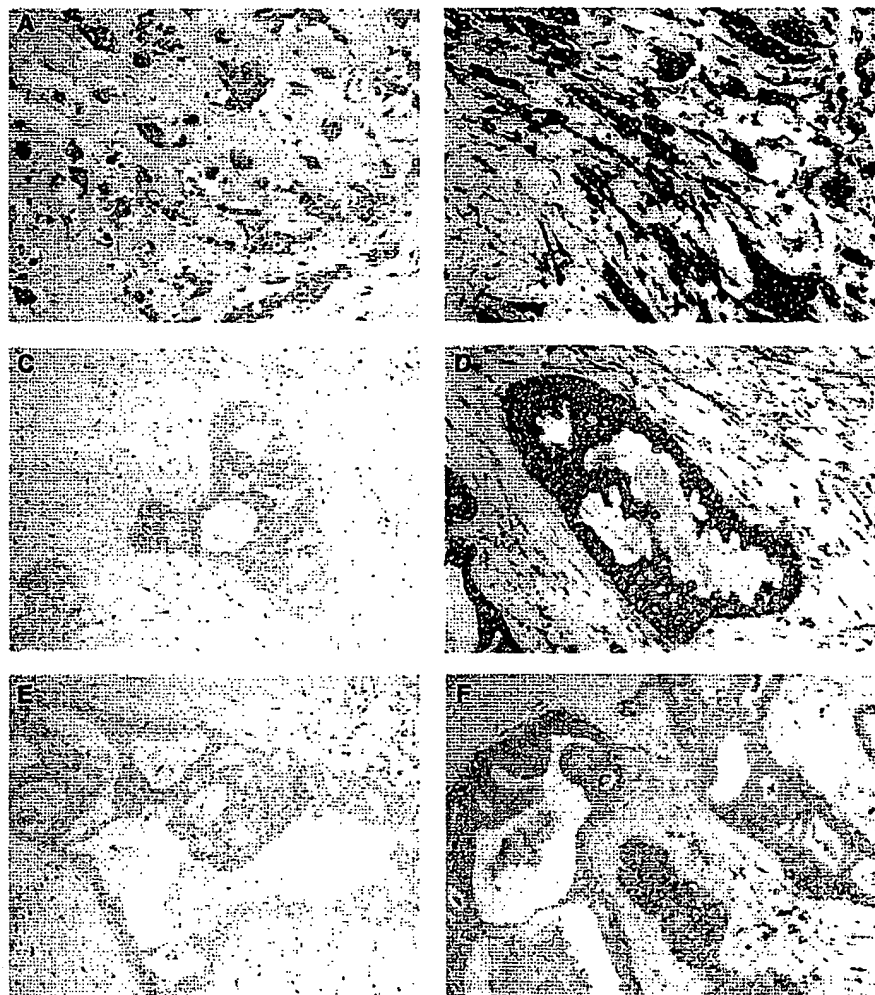


Figure 2 Immunohistochemical staining and *in situ* hybridisation for *c-Myc* of three sets of invasive ductal carcinoma. (A, C and E) High (3+), intermediate (2+) and low (1+) level of staining by immunohistochemistry for *c-Myc*. (B, D and F) High (3+), intermediate (2+) and low (1+) level of staining by *in situ* hybridisation.

DISCUSSION

Although there have been many reports on *c-myc* amplification in human breast cancer (Liao and Dickson, 2000), there are only two published studies involving application of the FISH technique to unfixed, frozen sections (Persons *et al.*, 1997; Visscher *et al.*, 1997), and one prior study using FISH on an archival human tissue microarray (Schraml *et al.*, 1999). Another recent study applied FISH to evaluate *c-myc* amplification in ductal carcinoma *in situ*

(DCIS) (Aulmann *et al.*, 2002). Using the FISH technique on formalin-fixed, paraffin-embedded sections, we now show that 70% of high-grade breast cancer samples bear *c-myc* gene copy amplifications. Interestingly, the above-mentioned study, using FISH and focusing on DCIS, detected amplification of *c-myc* in only 20% of cases, but found a correlation of *c-myc* with increased tumour size and proliferation (Aulmann *et al.*, 2002).

The level of amplification of *c-myc* in our study ranged between one and four additional copies of the gene; the majority (84%) of

the cases with the gene amplification gained only one to two copies, also consistent with FISH data reported for c-myc copy amplification in human metastatic prostate carcinoma tissues (Jenkins *et al*, 1997). The relationship between the level of c-myc gene copy amplification and the level its increased mRNA expression has been examined previously in breast cancer cell lines (Hyman *et al*, 2002). In general, it has been concluded that the two scores coordinate for c-myc, as is the case for many breast cancer genes. However, only 44% of the highly amplified genes, in general, showed increased RNA expression, and only 10.5% of the highly overexpressed genes were gene copy-amplified in the cell line study (Hyman *et al*, 2002). Another analysis was conducted to study of relationships between gene amplification and expression of 6095 genes in 37 intermediate grade human breast tumours. This study demonstrated that 62% of the highly amplified genes also showed elevated expression; overall, a two-fold change in DNA copy number was associated with a 1.5-fold change in mRNA levels. Overall, 12% of the variation in gene expression in the breast tumours studied was associated with gene copy number variation (Pollack *et al*, 2002). Further study of additional human breast tumours, at precisely defined grades and stages, will be necessary in order to more fully define the relationships between DNA copy numbers and expression of genes. The studies we report here indicate higher levels of c-Myc gene amplification and expression, than other previous reports in breast cancer. We believe that this is probably the result of our analysis of individual tumour cells in a well-defined set of high-grade breast tumours. Prior c-Myc expression and amplification microarray studies used tumour specimens which contain normal stromal components,

Table 3 Correlations between c-myc gene copy number (FISH) mRNA expression (ISH)

	FISH	
	Low	High
(A) ISH (% cells)		
Low	1	3
High	19	18
		$P = 0.0067$
(B) ISH (intensity)		
Low	2	5
High	18	16
		$P = 0.0006$

Serial sections of high-grade human breast carcinomas were scored for c-myc gene copy number (FISH, Table 1) and mRNA expression (ISH, Table 2). In (A), a positive correlation ($P = 0.0067$) was observed between tumour samples with a high percentage of cells demonstrating mRNA expression and a high c-myc gene copy number. A score of 2 or higher was classified as high on ISH, and a score of median or greater was categorised as high on FISH. In (B), a positive correlation ($P = 0.0006$) was shown between a high level of intensity for c-Myc RNA expression and a high c-myc gene copy number. Note that a pairwise comparison of FISH and ISH was not possible for all cases, due to incomplete overlap of cases analysed with each assay.

Table 4 c-Myc immunohistochemistry (IHC) results

Staining intensity	0	1	2	3	Percent positivity	1	2	3	4
Number of tumour samples in each category	15	13	20	3	Number of tumour samples in each category	2	2	1	29

In all, 51 high-grade human breast carcinomas were analysed to determine the relationships between c-Myc protein expression and c-myc gene *in situ* hybridisation results. Data are shown in two ways in the above table. First, overall staining intensity of c-Myc-positive cells was scored as 0, 1, 2, 3 (low to high), and the number of tumour samples at each level of staining is indicated on the line below. Next, in a random subset of these cases, the percentage of tumour cells staining was scored as 0, 1, 2, 3, 4 (low to high % as discussed in Materials and methods). The number of tumours at each level of percent cell positivity for c-Myc is indicated on the line below.

potentially underestimating amplification and expression levels of the invasive tumour components (Pollack *et al*, 2002).

Our study reports a percentage of tumours gene amplified for c-myc (using FISH in high-grade tumours) that is much higher than the average figure (15.5%) reported in the literature (Isola *et al*, 2002). Most of the prior studies have employed the relatively insensitive Southern blot technique, and were reviewed in a recent meta-analysis (Deming *et al*, 2000). Consistent with this prior literature background, a recent study of 94 lobular and ductal breast cancers assessed amplification of c-myc by using a semiquantitative PCR assay and protein expression, with

Table 5 Nuclear/cytoplasmic localisation of c-Myc comparing normal and invasive cells

Normal cells (frequency percent)	Invasive cells (frequency percent)		Total
(A) Nuclear localisation			
	-	+	
-	28	4	32
+	14	8	22
Total	42	12	54
(B) Cytoplasmic localisation			
	-	+	
-	12	13	25
+	9	20	21
Total	21	33	54

In all, 54 pairs (normal vs invasive) of tissues were analysed to answer the questions of (1) whether positivity of nuclear cells in normal tissues is different from that in invasive cells, and similarly (2) whether positivity of cytoplasmic cells in normal tissues is different from that in invasive cells. The data are summarised in the above contingency tables. In all, 22 normal cell specimens were positive for c-Myc staining (40.71%), compared to 12 specimens (22.2%) in invasive cells. The difference is statistically significant ($P = 0.01$) by McNemar's χ^2 test (two-sided).

Table 6 Correlation between c-Myc protein expression (IHC) and c-myc gene copy number (FISH)

IHC (% cells)	FISH	
	Low	High
Low	3	0
High	10	15
		$P = 0.0016$

Consecutive serial sections of high-grade human breast tumours were scored for c-myc gene copy number or protein expression, by immunohistochemistry (IHC). IHC scores were defined in the Materials and methods section. Data were analysed for correlations between the results. A highly significant correlation was observed between high c-Myc protein expression (IHC) between percent cells positive and high c-myc gene amplification (FISH). $P = 0.0016$ from two-sided McNemar's test. Note that for 15 cases, no staining for c-Myc could be detected; these negative cases were not included in the correlation presented, above.

densitometry, after Western blot. These data showed c-myc gene amplification in 21% of tumours (Jenkins *et al*, 1997), using assays not based on *in situ* discrimination of tumour vs nontumour cells. The lower frequency of c-myc in this prior study is in contrast with the data we present here, and could be the result of the higher sensitivity and precision of the FISH and immunohistochemical methods, as distinct from quantitative PCR and Western blot densitometry. In addition, the 70% of amplified tumours in our study is also much higher than the 12% reported by Schraml *et al* (1999), using a c-myc FISH test on a tissue microarray. This large difference may be because the arrays are prepared from cores of paraffin-embedded tissue, as small as 0.6 mm in diameter which may contain too few tumour cells for complete analysis of amplification of a gene, such as c-myc. c-myc is known to be quite heterogeneous in its gene amplification within individual tumours (in contrast to *HER2/neu*, for example) (Persons *et al*, 1997).

Most previous reports on the expression of c-myc mRNA have utilised Northern blot, dot blot or PCR-based approaches, while just a few involved *in situ* hybridisation, which were primarily performed on frozen tissue sections (Liao and Dickson, 2000). Normal breast tissue is dominated by adipose cells, differing greatly from tumour tissue in its epithelial cellularity. Thus, normal and tumour tissues may not be rigorously compared by techniques involving RNA extraction from total tissue. Therefore, conclusions such as 'increased expression' may be more difficult to make from studies with Northern blot, dot blot and PCR-based techniques that require RNA extraction from tissues that have not been fastidiously micro-dissected for selection of tumour cells. Using a more sensitive, nonradioactive *in situ* hybridisation (ISH) approach on formalin-fixed, paraffin-embedded sections, we report herein high expression of c-myc mRNA in 92% of high-grade breast carcinomas. This figure is much higher than the recently reported data (22%), obtained by using a real-time RT-PCR method (Bieche *et al*, 1999). Dilution of the RNA from epithelium by the RNA from adipose in normal breast tissue in this latest prior report may be one of the possible explanations for this large difference.

In conclusion, the present study shows that approximately 70, 92 and 70% of biopsies of untreated high-grade breast cancer exhibit c-myc gene amplification, mRNA overexpression and protein overexpression, respectively. In most cases (84%), with gene copy

amplification, the c-myc gene gains one to two additional copies. c-myc gene amplification was significantly associated with expression of its mRNA (both by intensity in invasive cells and by percentage positivity in invasive cells), and with expression of its protein (by percentage positivity in invasive cells). However, our data were also consistent with the prior literature on c-Myc (reviewed in Nass and Dickson, 1997; Liao and Dickson, 2000), indicating complex transcriptional, post transcriptional, translational and post-translational control of c-Myc expression *in vitro*. Specifically, in Table 5 we observed that in 40% of the high-grade tumours tested, c-Myc protein was expressed at high levels, despite a lack of its gene amplification.

It will be interesting to analyse lower grade tumours and premalignant lesions, with the same measurement tools, to determine if this c-myc amplification pattern is different, comparing different steps in onset and progression of the disease. Specifically, prior studies in fibroblasts and in human mammary epithelial cells (Liao *et al*, 1998, 2000a, b) have demonstrated that only a subtle deregulation of expression of c-Myc is sufficient to allow genomic instability. These prior cell biologic findings raise the question of whether c-Myc protein expression precedes or follows its gene amplification during the course of the natural history of breast cancer. It will also be interesting for future studies of lower grade breast cancers and premalignant lesions to determine whether there is evidence of nuclear exclusion of c-Myc protein. Indeed, nuclear exclusion of c-Myc in high-grade tumours could serve to attenuate its functions in later stages of disease progression (Liao and Dickson, 2000).

ACKNOWLEDGEMENTS

We would like to express our gratitude to Mr G Veytsman and H-K Cao for scoring FISH data, Mrs S Constable for helping in tissue sectioning. This work was supported by NIH Grants RO1 CA72460 and AG1496 to RB Dickson and by a pilot grant from the Georgetown University ACS Institutional Research Grant to JK Blancato. This work was supported by NIH Grants RO1 CA72460 and AG1496 to RB Dickson and by an ACS pilot grant (from the GU Institutional ACS Research Grant) to JK Blancato.

REFERENCES

- Augenlich L, Wadler S, Gorner G, Richards C, Ryan L, Multani A, Pathak S, Benson A, Haller D, Heerdt B (1997) Low-level c-myc amplification in human colonic carcinoma cell lines and tumors: a frequent, p53-independent mutation associated with improved outcome in a randomized multi-institutional trial. *Cancer Res* 57: 1769–1775
- Aulmann S, Bentz M, Sinn HP (2002) C-myc oncogene amplification in ductal carcinoma *in situ* of the breast. *Breast Cancer Res Treat* 74: 25–31
- Berns E, Klijn J, Van Putten W, van Staveren I, Portengen H, Foekens J (1992) c-myc amplification is a better prognostic factor than *HER2/neu* amplification in primary breast cancer. *Cancer Res* 52: 1107–1113
- Bieche I, Laurendeau I, Tozlu S, Olivi M, Vidaud D, Lidereau R, Vidaud M (1999) Quantitation of MYC gene expression in sporadic breast tumors with a real-time transcription-PCR assay. *Cancer Res* 59: 2759–2765
- Chana J, Cree I, Foss A, Hungerford J, Wilson G (1998a) The prognostic significance of c-myc oncogene expression in uveal melanoma. *Melanoma Res* 8: 139–144
- Chana J, Grover R, Wilson G, Hudson D, Forders M, Sanders R, Grobbelaar A (1998b) The clinical significance of c-myc oncogene expression in melanomas of the scalp. *Br J Plast Surg* 51: 191–194
- Chana J, Wilson G, Cree I, Alexander R, Myatt N, Neale M, Foss A, Hungerford J (1999) c-myc, p53, and bcl-2 expression and clinical outcome in uveal melanoma. *Br J Ophthalmol* 83: 110–114
- Deming S, Nass S, Dickson R, Trock B (2000) c-myc amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. *Br J Cancer* 83: 1688–1695
- Diebold J, Suchy B, Baretton G, Blasenbren S, Meier W, Schmidt M, Rabes H, Lohrs U (1996) DNA ploidy and MYC DNA amplification in ovarian carcinomas: correlation with p53 and bcl-2 expression, proliferative activity and prognosis. *Virchows Arch* 429: 221–227
- Grover R, Chana J, Grobbelaar A, Hudson D, Forders M, Wilson G, Sanders R (1999) Measurement of c-myc oncogene expression provides an accurate prognostic marker for acral lentiginous melanoma. *Br J Plast Surg* 52: 122–126
- Grover R, Ross D, Wilson G, Sanders R (1997) Measurement of c-myc oncoprotein provides an independent prognostic marker for regional metastatic melanoma. *Br J Plast Surg* 50: 478–482
- Hymen E, Kauraniemi P, Hautaniemi S, Wolf M, Mousset S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahoul A, Kallioniemi O, Kallioniemi A (2002) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* 62: 6240–6245
- Isola J, Chu L, De Vries S, Matsumura K, Chew K, Ljung B, Waldman F (1999) Genetic alterations in ERBB2-amplified breast carcinomas. *Clin Cancer Res* 5: 4140–4145
- Jenkins R, Qian J, Lieber M, Bostwick D (1997) Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic

- prostatic carcinoma by fluorescence *in situ* hybridization. *Cancer Res* 57: 524-531
- Liao D, Dickson R (2000) c-Myc in breast cancer. *Endocrine-Relat Cancer* 7: 143-164
- Liao D, Hou X, Bai S, Li S, Li J (2000a) Unusual deregulation of cell cycle components in early and frank estrogen-induced renal neoplasias in the Syrian hamster. *Carcinogenesis* 21: 2167-2173
- Liao D, Natarajan G, Deming S, Jamerson M, Johnson M, Chepko G, Dickson R (2000b) Cell cycle basis for the onset and progression of c-Myc-induced, TGF α -enhanced mouse mammary gland carcinogenesis. *Oncogene* 19: 1307-1317
- Liao D, Pantazis C, Hou X, Li S (1998) Promotion of estrogen-induced mammary gland carcinogenesis by androgen in the male Noble rat: probable mediation by steroid receptors. *Carcinogenesis* 19: 2173-2180
- Marcu K, Bossone S, Patel A (1992) Myc function and regulation. *Annu Rev Biochem* 61: 809-860
- Melhem M, Meisler A, Finley G, Bryce W, Jones M, Tribby I, Pipas J, Koski R (1992) Distribution of cells expressing Myc proteins in human colorectal epithelium, polyps, and malignant tumors. *Cancer Res* 52: 5853-5864
- Nass S, Dickson R (1997) Defining a role for c-Myc in breast tumorigenesis. *Breast Cancer Res Treat* 44: 1-22
- Nesbit C, Tersak J, Prochownik E (1999) MYC oncogenes and human neoplastic disease. *Oncogene* 18: 3004-3016
- Persons D, Borelli K, Hsu P (1997) Quantitation of *HER-2/neu* and c-myc gene amplification in breast carcinoma using fluorescence *in situ* hybridization. *Mod Pathol* 10: 720-727
- Pietilainen T, Lipponen P, Aaltomaa S, Eskelinen M, Kosma V, Syrjanen K (1995) Expression of c-myc proteins in breast cancer as related to established prognostic factors and survival. *Anticancer Res* 15: 959-964
- Polaczar S, Hey N, Stephenson T, Hill A (1989) c-myc oncogene product p62^{c-myc} in ovarian mucinous neoplasms: immunohistochemical study correlated with malignancy. *J Clin Pathol* 42: 148-152
- Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* 99: 12963-12968
- Sartelet S, Grossi L, Pasquier D, Conbaret V, Bouvier R, Ranchere D, Plantaz D, Munzer M, Philip T, Birembaut P, Zahn J, Bergeron C, Gaillard D, Pasquier B (2002) Detection of N-myc amplification by FISH in immature areas of fixed neuroblastomas: more efficient than Southern blot/PCR. *J Pathol* 198: 83-91
- Sasano H, Nagura H, Silverberg S (1992) Immunolocalization of c-myc oncoprotein in mucinous and serous adenocarcinomas of the ovary. *Hum Pathol* 23: 491-495
- Sato H, Tsuchiya A, Abe R (1995) Correlation between c-myc protein expression and phases of the cell cycle in human colorectal carcinomas. *Fukushima J Med Sci* 41: 111-113
- Schraml P, Kononen J, Bubendorf L, Moch H, Bissig H, Noctito A, Mihatsch M, Kallioniemi O, Sauter G (1999) Tissue microarrays for gene amplification surveys in many different tumor types. *Clin Cancer Res* 5: 1966-1975
- Sikora K, Chan S, Evan G, Gabra H, Markham N, Stewart J, Watson J (1987) c-myc oncogene expression in colorectal cancer. *Cancer* 59: 1289-1295
- Sikora K, Evan G, Stewart J, Watson J (1985) Detection of the c-myc oncogene product in testicular cancer. *Br J Cancer* 52: 171-176
- Smith D, Goh H-S (1996) Overexpression of c-myc proto-oncogene in colorectal carcinoma is associated with a reduced mortality that is abrogated by point mutation of the p53 tumor suppressor gene. *Clin Cancer Res* 2: 1049-1053
- Sundaresan V, Forgacs I, Wight D, Wilson B, Evan G, Watson J (1987) Abnormal distribution of c-myc oncogene product in familial adenomatous polyposis. *J Clin Pathol* 40: 1274-1281
- Visca P, Alo P, Del Nonno F, Bott C, Trombetta G, Marandino F, Filippi S, Di Tondo U, Donnorso R (1999) Immunohistochemical expression of fatty acid synthase, apoptotic-regulating genes, proliferating factors, and ras, protein product in colorectal adenomas, carcinomas, and adjacent nonneoplastic mucosa. *Clin Cancer Res* 5: 4111-4118
- Visscher D, Wallis T, Awussah S, Mohamed A, Crissman J (1997) Evaluation of MYC and chromosome 8 copy number in breast carcinoma by interphase cytogenetics. *Genes, Chromosomes, Cancer* 6: 389-404
- Voravud N, Foster C, Gilbertson J, Sikora K, Waxman J (1989) Oncogene expression in cholangiocarcinoma and in normal hepatic development. *Hum Pathol* 20: 1163-1168
- Watson J, Stewart J, Evan G, Ritson A, Sikora K (1986) The clinical significance of low cytometric c-myc oncoprotein in testicular cancer. *Br J Cancer* 53: 331-337
- Williams A, Piris J, Wyllie A (1990) Immunohistochemical demonstration of altered intracellular localization of the c-myc oncogene product in human colorectal neoplasms. *J Pathol* 160: 287-293
- Yuen M, Wu P, Lau J, Lai C (2001) Expression of c-Myc, c-Fos, and c-Jun in hepatocellular carcinoma. *Cancer* 91: 106-112
- Zar JH (1974) *Biostatistical Analysis*. Prentice-Hall: Englewood Cliffs, NJ

MOLECULAR BIOLOGY OF
THE CELL
THIRD EDITION

Text Editor: Miranda Robertson
Managing Editor: Ruth Adams
Illustrator: Nigel Orme
Molecular Model Drawings: Kate Hesketh-Moore
Director of Electronic Publishing: John M-Roblin
Computer Specialist: Chuck Bartelt
Disk Preparation: Carol Winter
Copy Editor: Shirley M. Cobert
Production Editor: Douglas Goertzen
Production Coordinator: Perry Bessas
Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge.

Julian Lewis received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Molecular biology of the cell / Bruce Alberts . . . [et al.].—3rd ed.
p. cm.

Includes bibliographical references and index.

ISBN 0-8153-1619-4 (hard cover).—ISBN 0-8153-1620-8 (pbk.)

1. Cytology. 2. Molecular biology. I. Alberts, Bruce.

[DNLM: 1. Cells. 2. Molecular Biology. QH 581.2 M718 1994]

QH581.2.M64 1994

574.87—dc20

DNLM/DLC

for Library of Congress

93-45907
CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

15 14 13 12 10 9 8 7

Front cover: The photograph shows a rat nerve cell in culture. It is labeled (*yellow*) with a fluorescent antibody that stains its cell body and dendritic processes. Nerve terminals (*green*) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gavin Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoC author Bruce Alberts and famous mountaineer guide Mugs Stump (1940–1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Olivier.)

extracts. If these minor cell proteins differ among cells to the same extent as the more abundant proteins, as is commonly assumed, only a small number of protein differences (perhaps several hundred) suffice to create very large differences in cell morphology and behavior.

A Cell Can Change the Expression of Its Genes in Response to External Signals³

Most of the specialized cells in a multicellular organism are capable of altering their patterns of gene expression in response to extracellular cues. If a liver cell is exposed to a glucocorticoid hormone, for example, the production of several specific proteins is dramatically increased. Glucocorticoids are released during periods of starvation or intense exercise and signal the liver to increase the production of glucose from amino acids and other small molecules; the set of proteins whose production is induced includes enzymes such as tyrosine aminotransferase, which helps to convert tyrosine to glucose. When the hormone is no longer present, the production of these proteins drops to its normal level.

Other cell types respond to glucocorticoids in different ways. In fat cells, for example, the production of tyrosine aminotransferase is reduced, while some other cell types do not respond to glucocorticoids at all. These examples illustrate a general feature of cell specialization—different cell types often respond in different ways to the same extracellular signal. Underlying this specialization are features that do not change, which give each cell type its permanently distinctive character. These features reflect the persistent expression of different sets of genes.

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein⁴

If differences between the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? There are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the primary RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytoplasm (**RNA transport control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 9-2).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9-2, only transcriptional control ensures that no superfluous intermediates are synthesized. In the

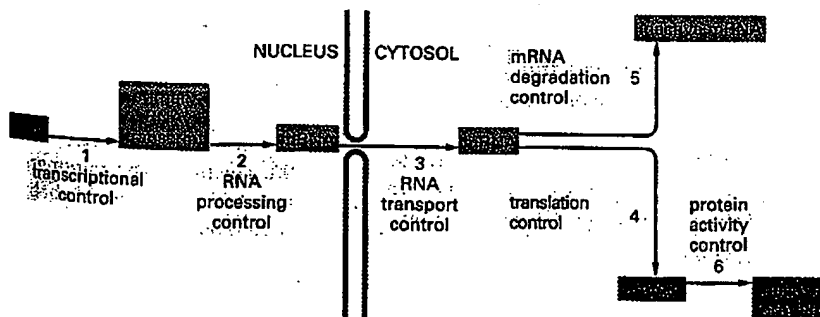


Figure 9-2 Six steps at which eucaryote gene expression can be controlled. Only controls that operate at steps 1 through 5 are discussed in this chapter. The regulation of protein activity (step 6) is discussed in Chapter 5; this includes reversible activation or inactivation by protein phosphorylation as well as irreversible inactivation by proteolytic degradation.

following sections we discuss the DNA and protein components that regulate the initiation of gene transcription. We return at the end of the chapter to the other ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-binding Motifs in Gene Regulatory Proteins⁵

How does a cell determine which of its thousands of genes to transcribe? As discussed in Chapter 8, the transcription of each gene is controlled by a regulatory region of DNA near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Other regulatory regions are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices consist of two fundamental types of components: (1) short stretches of DNA of defined sequence and (2) *gene regulatory proteins* that recognize and bind to them.

We begin our discussion of gene regulatory proteins by describing how these proteins were discovered.

Gene Regulatory Proteins Were Discovered Using Bacterial Genetics⁶

Genetic analyses in bacteria carried out in the 1950s provided the first evidence of the existence of **gene regulatory proteins** that turn specific sets of genes on or off. One of these regulators, the *lambda repressor*, is encoded by a bacterial virus, *bacteriophage lambda*. The repressor shuts off the viral genes that code for the protein components of new virus particles and thereby enables the viral genome to remain a silent passenger in the bacterial chromosome, multiplying with the bacterium when conditions are favorable for bacterial growth (see Figure 6-80). The lambda repressor was among the first gene regulatory proteins to be characterized, and it remains one of the best understood, as we discuss later. Other bacterial regulators respond to nutritional conditions by shutting off genes encoding specific sets of metabolic enzymes when they are not needed. The *lac repressor*, for example, the first of these bacterial proteins to be recognized, turns off the production of the proteins responsible for lactose metabolism when this sugar is absent from the medium.

The first step toward understanding gene regulation was the isolation of mutant strains of bacteria and bacteriophage lambda that were unable to shut off specific sets of genes. It was proposed at the time, and later proved, that most of these mutants were deficient in proteins acting as specific repressors for these sets of genes. Because these proteins, like most gene regulatory proteins, are present in small quantities, it was difficult and time-consuming to isolate them. They were eventually purified by fractionating cell extracts on a series of standard chromatography columns (see pp. 166-169). Once isolated, the proteins were shown to bind to specific DNA sequences close to the genes that they

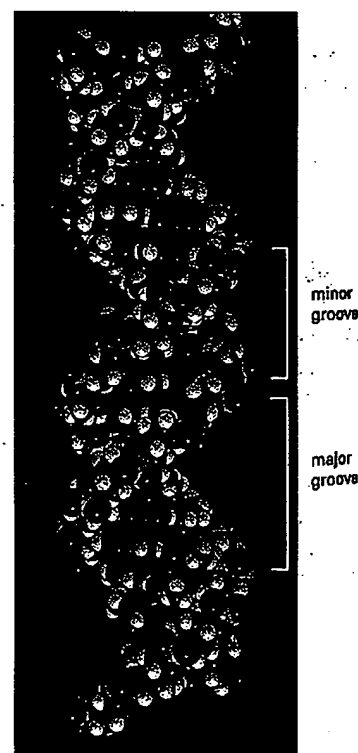


Figure 9-3 Double-helical structure of DNA. The major and minor grooves on the outside of the double helix are indicated. The atoms are colored as follows: carbon, dark blue; nitrogen, light blue; hydrogen, white; oxygen, red; phosphorus, yellow.

regulate
by a cor
experim

The O

As disc
double l
otide se
these pi
pairs in
and anc
ded wit
without
at the s
bond d
to reco
major g
(Figure
for gro
Alt
most ir
only or
double

DNA.

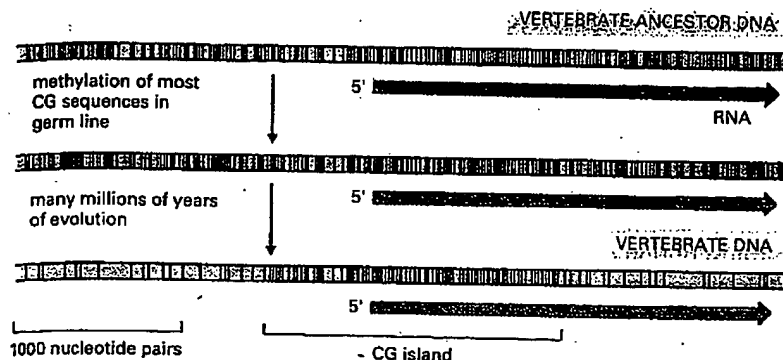


Figure 9-71 A mechanism to explain both the marked deficiency of CG sequences and the presence of CG islands in vertebrate genomes. A black line marks the location of an unmethylated CG dinucleotide in the DNA sequence, while a red line marks the location of a methylated CG dinucleotide.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides, endowing the cell with a memory of its developmental history. Prokaryotes and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms, some of which may be relevant to the creation of specialized cell types in higher eucaryotes. One such mechanism involves a competitive interaction between two (or more) gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory.

In eucaryotes gene transcription is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be expressed in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also utilized by eucaryotic cells to regulate gene expression. In vertebrates DNA methylation also plays a part, mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms.

Posttranscriptional Controls

Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made. Although these posttranscriptional controls, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than transcriptional control, for many genes they are crucial. It seems that every step in gene expression that could be controlled in principle is likely to be regulated under some circumstances for some genes.

We consider the varieties of posttranscriptional regulation in temporal order, according to the sequence of events that might be experienced by an RNA molecule after its transcription has begun (Figure 9-72).

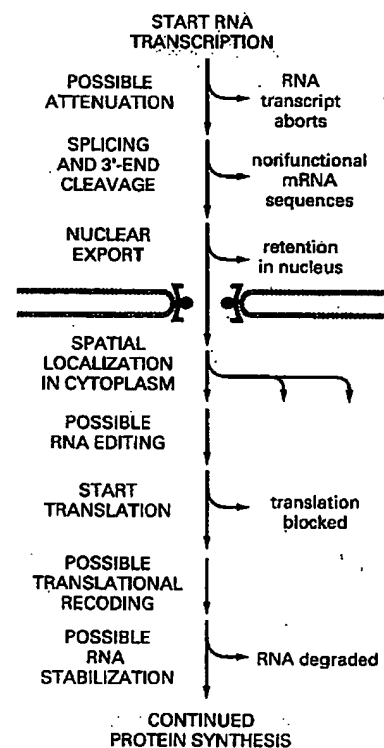


Figure 9-72 Possible post-transcriptional controls on gene expression. Only a few of these controls are likely to be used for any one gene.

MOLECULAR BIOLOGY OF
THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

GS Garland Science
Taylor & Francis Group

Garland

Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dileria
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

Cell Biology Interactive

Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

© 2002 by Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.
© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any format in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Molecular biology of the cell / Bruce Alberts ... [et al.]. -- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology.]
QH581.2 .M64 2002
571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group,
29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Front cover Human Genome: Reprinted by permission from *Nature*, International Human Genome Sequencing Consortium, 409:860–921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Ewan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

Back cover In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original. *Drosophila*, *Arabidopsis*, Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all." (Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1964; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s, © 1986 Delta Haze Corporation all rights reserved, used by permission; Albert L. Lehninger, (unidentified photographer) courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Poussin, courtesy of ArtToday.com; Barbara McClintock, © David Micklos, 1983; Andrei Sakharov, courtesy of Elena Bonner; Frederick Sanger, © The Nobel Foundation, 1958.)

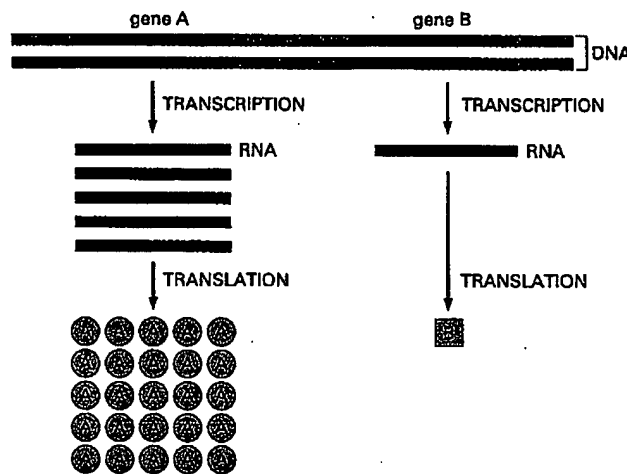


Figure 6-3 Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6-3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most obviously by controlling the production of its RNA.

Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence. The information in RNA, although copied into another chemical form, is still written in essentially the same language as it is in DNA—the language of a nucleotide sequence. Hence the name **transcription**.

Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds (Figure 6-4). It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribonucleic acid*) rather than deoxyribose; (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 6-5), the complementary base-pairing properties described for DNA in Chapters 4 and 5 apply also to RNA (in RNA, G pairs with C, and A pairs with U). It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.

Despite these small chemical differences, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is single-stranded. RNA chains therefore fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (Figure 6-6). As we see later in this chapter, the ability to fold into complex three-dimensional shapes allows some RNA molecules to have structural and catalytic functions.

Transcription Produces RNA Complementary to One Strand of DNA

All of the RNA in a cell is made by DNA transcription, a process that has certain similarities to the process of DNA replication discussed in Chapter 5.

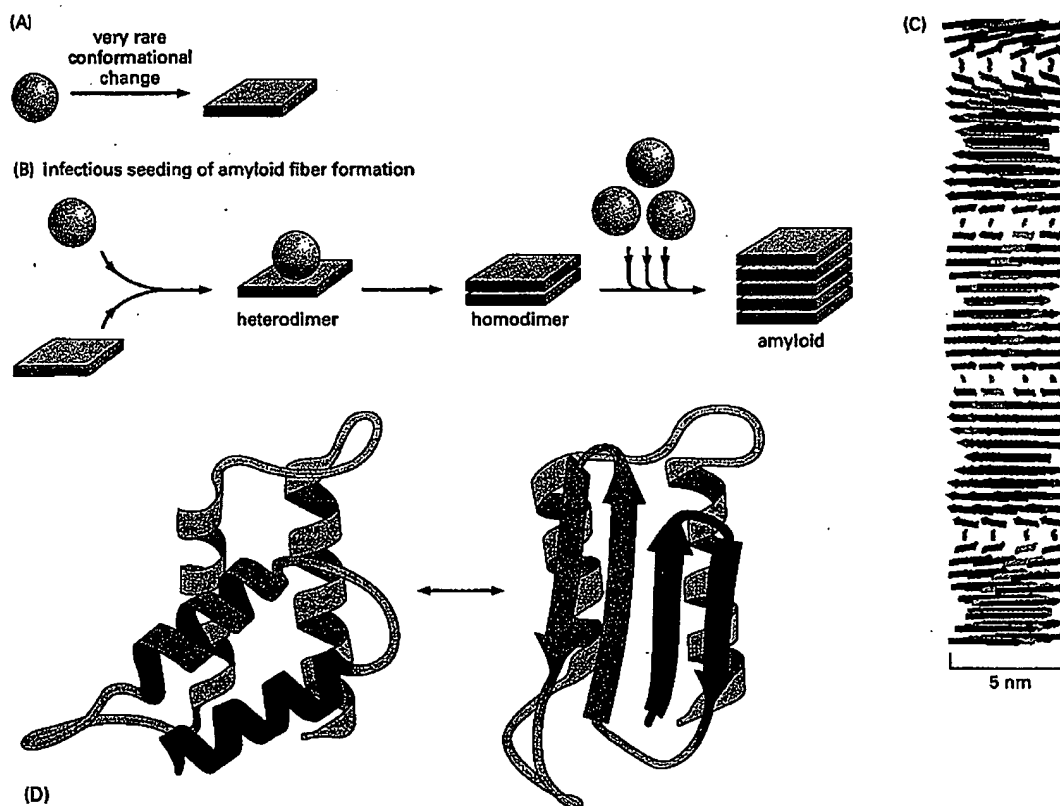


Figure 6-89 Protein aggregates that cause human disease. (A) Schematic illustration of the type of conformational change in a protein that produces material for a cross-beta filament. (B) Diagram illustrating the self-infectious nature of the protein aggregation that is central to prion diseases. PrP is highly unusual because the misfolded version of the protein, called PrP^{*}, induces the normal PrP protein it contacts to change its conformation, as shown. Most of the human diseases caused by protein aggregation are caused by the overproduction of a variant protein that is especially prone to aggregation, but because this structure is not infectious in this way, it cannot spread from one animal to another. (C) Drawing of a cross-beta filament, a common type of protease-resistant protein aggregate found in a variety of human neurological diseases. Because the hydrogen-bond interactions in a β sheet form between polypeptide backbone atoms (see Figure 3-9), a number of different abnormally folded proteins can produce this structure. (D) One of several possible models for the conversion of PrP to PrP^{*}, showing the likely change of two α -helices into four β -strands. Although the structure of the normal protein has been determined accurately, the structure of the infectious form is not yet known with certainty because the aggregation has prevented the use of standard structural techniques. (C, courtesy of Louise Serpell, adapted from M. Sunde et al., *J. Mol. Biol.* 273:729-739, 1997; D, adapted from S.B. Prusiner, *Trends Biochem. Sci.* 21:482-487, 1996.)

animals and humans. It can be dangerous to eat the tissues of animals that contain PrP^{*}, as witnessed most recently by the spread of BSE (commonly referred to as the "mad cow disease") from cattle to humans in Great Britain.

Fortunately, in the absence of PrP^{*}, PrP is extraordinarily difficult to convert to its abnormal form. Although very few proteins have the potential to misfold into an infectious conformation, a similar transformation has been discovered to be the cause of an otherwise mysterious "protein-only inheritance" observed in yeast cells.

There Are Many Steps From DNA to Protein

We have seen so far in this chapter that many different types of chemical reactions are required to produce a properly folded protein from the information contained in a gene (Figure 6-90). The final level of a properly folded protein in a cell therefore depends upon the efficiency with which each of the many steps is performed.

We discuss in Chapter 7 that cells have the ability to change the levels of their proteins according to their needs. In principle, any or all of the steps in Fig-

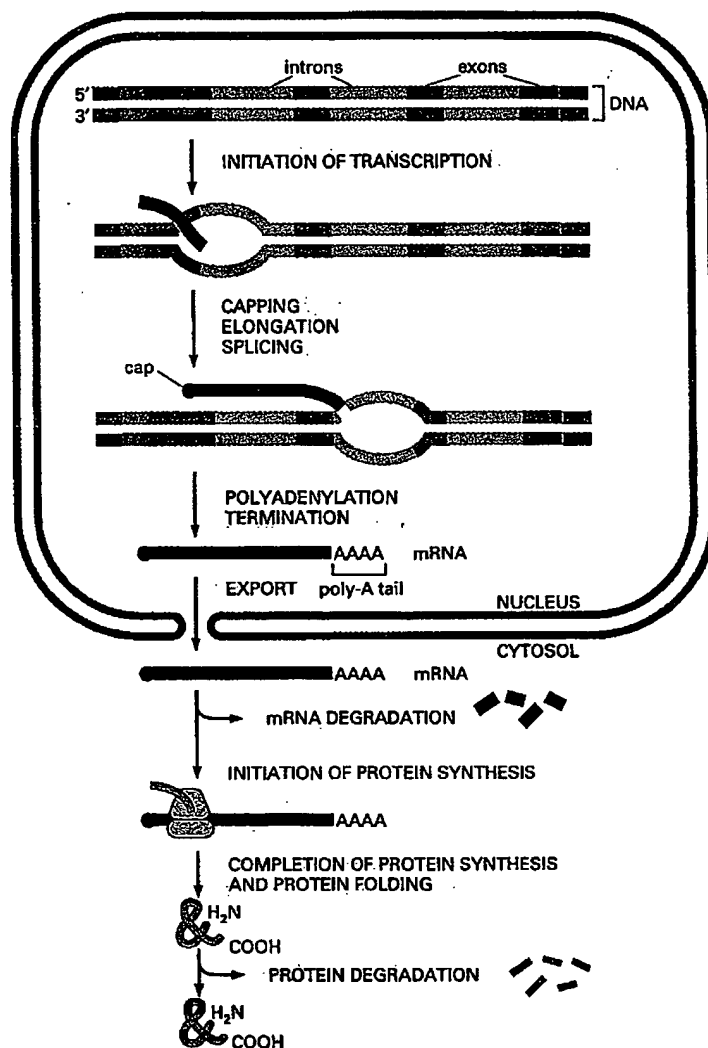


Figure 6-90 The production of a protein by a eucaryotic cell. The final level of each protein in a eucaryotic cell depends upon the efficiency of each step depicted.

ure 6-90) could be regulated by the cell for each individual protein. However, as we shall see in Chapter 7, the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes. This makes sense, inasmuch as the most efficient way to keep a gene from being expressed is to block the very first step—the transcription of its DNA sequence into an RNA molecule.

Summary

The translation of the nucleotide sequence of an mRNA molecule into protein takes place in the cytoplasm on a large ribonucleoprotein assembly called a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pair interactions, particular sets of three nucleotides in the mRNA (codons). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three according to the genetic code.

To initiate translation, a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and begin the elongation phase of protein synthesis. During this phase, aminoacyl tRNAs—each bearing a specific amino acid bind sequentially to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the C-terminal end of the growing polypeptide by means of a cycle of three sequential

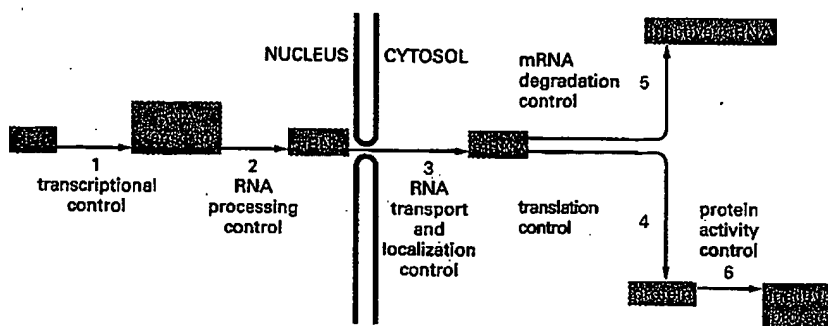


Figure 7-5 Six steps at which eucaryotic gene expression can be controlled. Controls that operate at steps 1 through 5 are discussed in this chapter. Step 6, the regulation of protein activity, includes reversible activation or inactivation by protein phosphorylation (discussed in Chapter 3) as well as irreversible inactivation by proteolytic degradation (discussed in Chapter 6).

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

If differences among the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? As we saw in the last chapter, there are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized (**RNA transport and localization control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, degrading, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 7-5).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 7-5, only transcriptional control ensures that the cell will not synthesize superfluous intermediates. In the following sections we discuss the DNA and protein components that perform this function by regulating the initiation of gene transcription. We shall return at the end of the chapter to the additional ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-BINDING MOTIFS IN GENE REGULATORY PROTEINS

How does a cell determine which of its thousands of genes to transcribe? As mentioned briefly in Chapters 4 and 6, the transcription of each gene is controlled by a regulatory region of DNA relatively near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Many others are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices

occur in the germ line, the cell lineage that gives rise to sperm or eggs. Most of the DNA in vertebrate germ cells is inactive and highly methylated. Over long periods of evolutionary time, the methylated CG sequences in these inactive regions have presumably been lost through spontaneous deamination events that were not properly repaired. However promoters of genes that remain active in the germ cell lineages (including most housekeeping genes) are kept unmethylated, and therefore spontaneous deaminations of Cs that occur within them can be accurately repaired. Such regions are preserved in modern day vertebrate cells as CG islands. In addition, any mutation of a CG sequence in the genome that destroyed the function or regulation of a gene in the adult would be selected against, and some CG islands are simply the result of a higher than normal density of critical CG sequences.

The mammalian genome contains an estimated 20,000 CG islands. Most of the islands mark the 5' ends of transcription units and thus, presumably, of genes. The presence of CG islands often provides a convenient way of identifying genes in the DNA sequences of vertebrate genomes.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character through many cell division cycles and even when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides. These features endow the cell with a memory of its developmental history. Bacteria and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms. One such mechanism involves a competitive interaction between two gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory. Negative feedback loops with programmed delays form the basis for cellular clocks.

In eucaryotes the transcription of a gene is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be active in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also used by eucaryotic cells to regulate gene expression. An especially dramatic case is the inactivation of an entire X chromosome in female mammals. In vertebrates DNA methylation also functions in gene regulation, being used mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms. DNA methylation also underlies the phenomenon of genomic imprinting in mammals, in which the expression of a gene depends on whether it was inherited from the mother or the father.

POSTTRANSCRIPTIONAL CONTROLS

In principle, every step required for the process of gene expression could be controlled. Indeed, one can find examples of each type of regulation, although any one gene is likely to use only a few of them. Controls on the initiation of gene transcription are the predominant form of regulation for most genes. But other controls can act later in the pathway from DNA to protein to modulate the amount of gene product that is made. Although these **posttranscriptional controls**, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than *transcriptional control*, for many genes they are crucial.

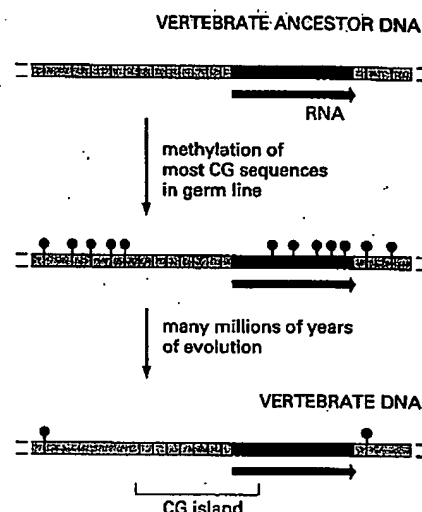


Figure 7-86 A mechanism to explain both the marked overall deficiency of CG sequences and their clustering into CG islands in vertebrate genomes. A black line marks the location of a CG dinucleotide in the DNA sequence, while a red "lollipop" indicates the presence of a methyl group on the CG dinucleotide. CG sequences that lie in regulatory sequences of genes that are transcribed in germ cells are unmethylated and therefore tend to be retained in evolution. Methylated CG sequences, on the other hand, tend to be lost through deamination of 5-methyl C to T, unless the CG sequence is critical for survival.

CHAPTER 29

Regulation of transcription

Genes VI (1997) CH 29, pp. 847-848.
Benjamin Lewin

The phenotypic differences that distinguish the various kinds of cells in a higher eukaryote are largely due to differences in the expression of genes that code for proteins, that is, those transcribed by RNA polymerase II. In principle, the expression of these genes might be regulated at any one of several stages. The concept of the "level of control" implies that gene expression is not necessarily an automatic process once it has begun. It could be regulated in a gene-specific way at any one of several sequential steps. We can distinguish (at least) five potential control points, forming the series:

Activation of gene structure
↓
Initiation of transcription
↓
Processing the transcript
↓
Transport to cytoplasm
↓
Translation of mRNA

The existence of the first step is implied by the discovery that genes may exist in either of two structural conditions. Relative to the state of most of the genome, genes are found in an "active" state in the cells in which they are expressed (see Chapter 27). The change of structure is distinct from the act of transcription, and indicates that the gene is "transcribable." This suggests that acquisition of the "active" structure must be the first step in gene expression.

Transcription of a gene in the active state is

controlled at the stage of initiation, that is, by the interaction of RNA polymerase with its promoter. This is now becoming susceptible to analysis in the *in vitro* systems (see Chapter 28). For most genes, this is a major control point; probably it is the most common level of regulation.

There is at present no evidence for control at subsequent stages of transcription in eukaryotic cells, for example, via antitermination mechanisms.

The primary transcript is modified by capping at the 5' end, and usually also by polyadenylation at the 3' end. Introns must be spliced out from the transcripts of interrupted genes. The mature RNA must be exported from the nucleus to the cytoplasm. Regulation of gene expression by selection of sequences at the level of nuclear RNA might involve any or all of these stages, but the one for which we have most evidence concerns changes in splicing: some genes are expressed by means of alternative splicing patterns whose regulation controls the type of protein product (see Chapter 30).

Finally, the translation of an mRNA in the cytoplasm can be specifically controlled. There is little evidence for the employment of this mechanism in adult somatic cells, but it does occur in some embryonic situations, as described in Chapter 7. The mechanism is presumed to involve the blocking of initiation of translation of some mRNAs by specific protein factors.

But having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear

that the overwhelming majority of regulatory events occur at the initiation of transcription. Regulation of tissue-specific gene transcription lies at the heart of eukaryotic differentiation; indeed, we see examples in Chapter 38 in which proteins that regulate embryonic development prove to be transcription factors. A regulatory transcription factor serves to provide

common control of a large number of target genes, and we seek to answer two questions about this mode of regulation: what identifies the common target genes to the transcription factor; and how is the activity of the transcription factor itself regulated in response to intrinsic or extrinsic signals?

Response elements identify genes under common regulation

The principle that emerges from characterizing groups of genes under common control is that *they share a promoter element that is recognized by a regulatory transcription factor*. An element that causes a gene to respond to such a factor is called a response element; examples are the HSE (heat shock response element), GRE (glucocorticoid response element), SRE (serum response element).

The properties of some inducible transcription factors and the elements that they recognize are summarized in Table 29.1. Response elements have the same general characteristics as upstream elements of promoters or enhancers. They contain short consensus sequences, and copies of the response elements found in different genes are closely related, but not necessarily identical. The region bound by the factor extends for a short distance on either side of

the consensus sequence. In promoters, the elements are not present at fixed distances from the startpoint, but are usually <200 bp upstream of it. The presence of a single element usually is sufficient to confer the regulatory response, but sometimes there are multiple copies.

Response elements may be located in promoters or in enhancers. Some types of elements are typically found in one rather than the other: usually an HSE is found in a promoter, while a GRE is found in an enhancer. We assume that all response elements function by the same general principle. *A gene is regulated by a sequence at the promoter or enhancer that is recognized by a specific protein. The protein functions as a transcription factor needed for RNA polymerase to initiate. Active protein is available only under conditions when the gene is to be expressed; its absence means that the promoter is not activated by this particular circuit.*

An example of a situation in which many genes are controlled by a single factor is provided by the heat shock response. This is common to a wide range of prokaryotes and eukaryotes and involves multiple controls of gene expression: an increase in temperature turns off transcription of some genes, turns on transcription of the heat shock genes, and causes changes in the translation of mRNAs. The control of the heat shock genes illustrates the differences between prokaryotic and eukaryotic modes of control. In bacteria, a new sigma factor is synthesized that directs RNA polymerase holoenzyme to recognize an alter-

Table 29.1 Inducible transcription factors bind to response elements that identify groups of promoters or enhancers subject to coordinate control.

Regulatory Agent	Module	Consensus	Factor
Heat shock	HSE	CNNGAANNTCGNG	HSTF
Glucocorticoid	GRE	TGGTACAAATGTTCT	Receptor
Phorbol ester	TRE	TGACTCA	AP1
Serum	SRE	CCATATTAGG	SRF

Research

Open Access

Prostate stem cell antigen (PSCA) expression in human prostate cancer tissues and its potential role in prostate carcinogenesis and progression of prostate cancer

Zhao Zhigang*¹ and Shen Wenlv²

Address: ¹Department of Urology, Shantou University Medical College, Shantou, Guangdong, China and ²Department of Urology, No 2. Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

Email: Zhao Zhigang* - zgzhao@163.com; Shen Wenlv - wshen99@hotmail.com

* Corresponding author

Published: 10 May 2004

Received: 30 March 2004

World Journal of Surgical Oncology 2004, 2:13

Accepted: 10 May 2004

This article is available from: <http://www.wjso.com/content/2/1/13>

© 2004 Zhigang and Wenlv; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Prostate stem cell antigen (PSCA) is a recently defined homologue of the Thy-1/Ly-6 family of glycosylphosphatidylinositol (GPI)-anchored cell surface antigens. The purpose of the present study was to examine the expression status of PSCA protein and mRNA in clinical specimens of human prostate cancer (Pca) and to validate it as a potential molecular target for diagnosis and treatment of Pca.

Materials and Methods: Immunohistochemical (IHC) and *in situ* hybridization (ISH) analyses of PSCA expression were simultaneously performed on paraffin-embedded sections from 20 benign prostatic hyperplasia (BPH), 20 prostatic intraepithelial neoplasm (PIN) and 48 prostate cancer (Pca) tissues, including 9 androgen-independent prostate cancers. The level of PSCA expression was semiquantitatively scored by assessing both the percentage and intensity of PSCA-positive staining cells in the specimens. Then compared PSCA expression between BPH, PIN and Pca tissues and analysed the correlations of PSCA expression level with pathological grade, clinical stage and progression to androgen-independence in Pca.

Results: In BPH and low grade PIN, PSCA protein and mRNA staining were weak or negative and less intense and uniform than that seen in HGPIN and Pca. There were moderate to strong PSCA protein and mRNA expression in 8 of 11 (72.7%) HGPIN and in 40 of 48 (83.4%) Pca specimens examined by IHC and ISH analyses, with statistical significance compared with BPH (20%) and low grade PIN (22.2%) samples ($p < 0.05$, respectively). The expression level of PSCA increased with high Gleason grade, advanced stage and progression to androgen-independence ($p < 0.05$, respectively). In addition, IHC and ISH staining showed a high degree of correlation between PSCA protein and mRNA overexpression.

Conclusions: Our data demonstrate that PSCA as a new cell surface marker is overexpressed by a majority of human Pca. PSCA expression correlates positively with adverse tumor characteristics, such as increasing pathological grade (poor cell differentiation), worsening clinical stage and androgen-independence, and speculatively with prostate carcinogenesis. PSCA protein overexpression results from upregulated transcription of PSCA mRNA. PSCA may have prognostic utility and may be a promising molecular target for diagnosis and treatment of Pca.

Introduction

Prostate cancer (Pca) is the second leading cause of cancer-related death in American men and is becoming a common cancer increasing in China. Despite recently great progress in the diagnosis and management of localized disease, there continues to be a need for new diagnostic markers that can accurately discriminate between indolent and aggressive variants of Pca. There also continues to be a need for the identification and characterization of potential new therapeutic targets on Pca cells. Current diagnostic and therapeutic modalities for recurrent and metastatic Pca have been limited by a lack of specific target antigens of Pca.

Although a number of prostate-specific genes have been identified (i.e. prostate specific antigen, prostatic acid phosphatase, glandular kallikrein 2), the majority of these are secreted proteins not ideally suited for many immunological strategies. So, the identification of new cell surface antigens is critical to the development of new diagnostic and therapeutic approaches to the management of Pca.

Reiter RE et al [1] reported the identification of prostate stem cell antigen (PSCA), a cell surface antigen that is predominantly prostate specific. The PSCA gene encodes a 123 amino acid glycoprotein, with 30% homology to stem cell antigen 2 (Sca 2). Like Sca-2, PSCA also belongs to a member of the Thy-1/Ly-6 family and is anchored by a glycosylphosphatidylinositol (GPI) linkage. mRNA *in situ* hybridization (ISH) localized PSCA expression in normal prostate to the basal cell epithelium, the putative stem cell compartment of prostatic epithelium, suggesting that PSCA may be a marker of prostate stem/progenitor cells.

In order to examine the status of PSCA protein and mRNA expression in human Pca and validate it as a potential diagnostic and therapeutic target for Pca, we used immunohistochemistry (IHC) and *in situ* hybridization (ISH) simultaneously, and conducted PSCA protein and mRNA expression analyses in paraffin-embedded tissue specimens of benign prostatic hyperplasia (BPH, n = 20), prostate intraepithelial neoplasm (PIN, n = 20) and prostate cancer (Pca, n = 48). Furthermore, we evaluated the possible correlation of PSCA expression level with Pca tumorigenesis, grade, stage and progression to androgen-independence.

Materials and methods

Tissue samples

All of the clinical tissue specimens studied herein were obtained from 80 patients of 57–84 years old by prostatectomy, transurethral resection of prostate (TURP) or biopsies. The patients were classified as 20 cases of BPH, 20 cases of PIN, 40 cases of primary Pca, including 9 patients

with recurrent Pca and a history of androgen ablation therapy (orchiectomy and/or hormonal therapy), who were referred to as androgen-independent prostate cancers. Eight specimens were harvested from these androgen-independent Pca patients prior to androgen ablation treatment. Each tissue sample was cut into two parts, one was fixed in 10% formalin for IHC and the other treated with 4% paraformaldehyde/0.1 M PBS PH 7.4 in 0.1% DEPC for 1 h for ISH analysis, and then embedded in paraffin. All paraffin blocks examined were then cut into 5 μ m sections and mounted on the glass slides specific for IHC and ISH respectively in the usual fashion. H&E-stained section of each Pca was evaluated and assigned a Gleason score by the experienced urological pathologist at our institution based on the criteria of Gleason score [2]. The Gleason sums are summarized in Table 1. Clinical staging was performed according to Jewett-whitmore-prout staging system, as shown in Table 2. In the category of PIN, we graded the specimens into two groups, i.e. low grade PIN (grade I – II) and high grade PIN (HG PIN, grade III) on the basis of literatures [3,4].

Immunohistochemical (IHC) analysis

Briefly, tissue sections were deparaffinized, dehydrated, and subjected to microwaving in 10 mmol/L citrate buffer, PH 6.0 (Boshide, Wuhan, China) in a 900 W oven for 5 min to induce epitope retrieval. Slides were allowed to cool at room temperature for 30 min. A primary mouse antibody specific to human PSCA (Boshide, Wuhan, China) with a 1:100 dilution was applied to incubate with the slides at room temperature for 2 h. Labeling was detected by sequentially adding biotinylated secondary antibodies and streptavidin-peroxidase, and localized using 3,3'-diaminobenzidine reaction. Sections were then counterstained with hematoxylin. Substitution of the primary antibody with phosphate-buffered-saline (PBS) served as a negative-staining control.

mRNA *in situ* hybridization (ISH)

Five- μ m-thick tissue sections were deparaffinized and dehydrated, then digested in pepsin solution (4 mg/ml in 3% citric acid) for 20 min at 37.5°C, and further processed for ISH. Digoxigenin-labeled sense and antisense human PSCA RNA probes (obtained from Boshide, Wuhan, China) were hybridized to the sections at 48°C overnight. The posthybridization wash with a high stringency was performed sequentially at 37°C in 2 \times standard saline citrate (SSC) for 10 min, in 0.5 \times SSC for 15 min and in 0.2 \times SSC for 30 min. The slides were then incubated to biotinylated mouse anti-digoxigenin antibody at 37.5°C for 1 h followed by washing in 1 \times PBS for 20 min at room temperature, and then to streptavidin-peroxidase at 37.5°C for 20 min followed by washing in 1 \times PBS for 15 min at room temperature. Subsequently, the slides were developed with diaminobenzidine and then coun-

Table 1: Correlation of PSCA expression with Gleason score

Gleason score	Intensity × frequency	
	0-6 (%)	9 (%)
2-4	5 (83)	1 (17)
5-7	19 (79)	5 (21)
8-10	5 (28)	13 (72)

Table 2: Correlation of PSCA expression with clinical stage

Tumor stage	Intensity × frequency	
	0-6 (%)	9 (%)
≤B	27 (67.5)	13 (32.5)
≥C	2 (25)	6 (75)

terstained with hematoxylin to localize the hybridization signals. Sections hybridized with the sense control probes routinely did not show any specific hybridization signal above background. All slides were hybridized with PBS to substitute for the probes as a negative control.

Scoring methods

To determine the correlation between the results of PSCA immunostaining and mRNA *in situ* hybridization, the same scoring manners are taken in the present study for PSCA protein staining by IHC and PSCA mRNA staining by ISH. Each slide was read and scored by two independently experienced urological pathologists using Olympus BX-41 light microscopes. The evaluation was done in a blinded fashion. For each section, five areas of similar grade were analyzed semiquantitatively for the fraction of cells staining. Fifty percent of specimens were randomly chosen and rescored to determine the degree of interobserver and intraobserver concordance. There was greater than 95% intra- and interobserver agreement.

The intensity of PSCA expression evaluated microscopically was graded on a scale of 0 to 3+ with 3 being the highest expression observed (0, no staining; 1+, mildly intense; 2+, moderately intense; 3+, severely intense). The staining density was quantified as the percentage of cells staining positive for PSCA with the primary antibody or hybridization probe, as follows: 0 = no staining; 1 = positive staining in <25% of the sample; 2 = positive staining in 25%–50% of the sample; 3 = positive staining in >50%

of the sample. Intensity score (0 to 3+) was multiplied by the density score (0–3) to give an overall score of 0–9 [1,5]. In this way, we were able to differentiate specimens that may have had focal areas of increased staining from those that had diffuse areas of increased staining [6]. The overall score for each specimen was then categorically assigned to one of the following groups: 0 score, negative expression; 1–2 scores, weak expression; 3–6 scores, moderate expression; 9 score, strong expression.

Statistical analysis

Intensity and density of PSCA protein and mRNA expression in BPH, PIN and Pca tissues were compared using the Chi-square and Student's *t*-test. Univariate associations between PSCA expression and Gleason score, clinical stage and progression to androgen-independence were calculated using Fisher's Exact Test. For all analyses, *p* < 0.05 was considered statistically significant.

Results

PSCA expression in BPH

In general, PSCA protein and mRNA were expressed weakly in individual samples of BPH. Some areas of prostate expressed weak levels (composite score 1–2), whereas other areas were completely negative (composite score 0). Four cases (20%) of BPH had moderate expression of PSCA protein and mRNA (composite score 4–6) by IHC and ISH. In 2/20 (10%) BPH specimens, PSCA mRNA expression was moderate (composite score 3–6), but PSCA protein expression was weak (composite score

2) in one and negative (composite score 0) in the other. PSCA expression was localized to the basal and secretory epithelial cells, and prostatic stroma was almost negative staining for PSCA protein and mRNA in all cases examined.

PSCA expression in PIN

In this study, we detected weak or negative expression of PSCA protein and mRNA (≤ 2 scores) in 7 of 9 (77.8%) low grade PIN and in 2 of 11 (18.2%) HGPIN, and moderate expression (3–6 scores) in the rest 2 low grade PIN and 5 of 11 (45.5%) HGPIN. One HGPIN with moderate PSCA mRNA expression (6 score) was found weak staining for PSCA protein (2 score) by IHC. Strong PSCA protein and mRNA expression (9 score) were detected in the remaining 3 of 11 (27.3%) HGPIN. There was a statistically significant difference of PSCA protein and mRNA expression levels observed between HGPIN and BPH ($p < 0.05$), but no statistical difference reached between low grade PIN and BPH ($p > 0.05$).

PSCA expression in Pca

In order to determine if PSCA protein and mRNA can be detected in prostate cancers and if PSCA expression levels are increased in malignant compared with benign glands, Forty-eight paraffin-embedded Pca specimens were analysed by IHC and ISH. It was shown that 19 of 48 (39.6%) Pca samples stained very strongly for PSCA protein and mRNA with a score of 9 and another 21 (43.8%) specimens displayed moderate staining with scores of 4–6 (Figure 1). In addition, 4 specimens with moderate to strong PSCA mRNA expression (scores of 4–9) had weak protein staining (a score of 2) by IHC analyses. Overall, Pca expressed a significantly higher level of PSCA protein and mRNA than any other specimen category in this study ($p < 0.05$, compared with BPH and PIN respectively). The result demonstrates that PSCA protein and mRNA are overexpressed by a majority of human Pca.

Correlation of PSCA expression with Gleason score in Pca

Using the semi-quantitative scoring method as described in Materials and Methods, we compared the expression level of PSCA protein and mRNA with Gleason grade of Pca, as shown in Table 1. Prostate adenocarcinomas were graded by Gleason score as 2–4 scores = well-differentiation, 5–7 scores = moderate-differentiation and 8–10 scores = poor-differentiation [7]. Seventy-two percent of Gleason scores 8–10 prostate cancers had very strong staining of PSCA compared to 21% with Gleason scores 5–7 and 17% with 2–4 respectively, demonstrating that poorly differentiated Pca had significantly stronger expression of PSCA protein and mRNA than moderately and well differentiated tumors ($p < 0.05$). As depicted in Figure 1, IHC and ISH analyses showed that PSCA protein and mRNA expression in several cases of poorly differen-

tiated Pca were particularly prominent, with more intense and uniform staining. The results indicate that PSCA expression increases significantly with higher tumor grade in human Pca.

Correlation of PSCA expression with clinical stage in Pca

With regards to PSCA expression in every stage of Pca, we showed the results in Table 2. Seventy-five percent of locally advanced and node positive cancers (i.e. C-D stages) expressed statistically high levels of PSCA versus 32.5% that were organ confined (i.e. A-B stages) ($p < 0.05$). The data demonstrate that PSCA expression increases significantly with advanced tumor stage in human Pca.

Correlation of PSCA expression with androgen-independent progression of Pca

All 9 specimens of androgen-independent prostate cancers stained positive for PSCA protein and mRNA. Eight specimens were obtained from patients managed prior to androgen ablation therapy. Seven of eight (87.5%) of these androgen-independent prostate cancers were in the strongest staining category (score = 9), compared with three out of eight (37.5%) of patients with androgen-dependent cancers ($p < 0.05$). The results demonstrate that PSCA expression increases significantly with progression to androgen-independence of human Pca.

It is evident from the results above that within a majority of human prostate cancers the level of PSCA protein and mRNA expression correlates significantly with increasing grade, worsening stage and progression to androgen-independence.

Correlation of PSCA immunostaining and mRNA in situ hybridization

In all 88 specimens surveyed herein, we compared the results of PSCA IHC staining with mRNA ISH analysis. Positive staining areas and its intensity and density scores evaluated by IHC were identical to those seen by ISH in 79 of 88 (89.8%) specimens (18/20 BPH, 19/20 PIN and 42/48 Pca respectively). Importantly, 27/27 samples with PSCA mRNA composite scores of 0–2, 32/36 samples with scores of 3–6 and 22/24 samples with a score of 9 also had PSCA protein expression scores of 0–2, 3–6 and 9 respectively. However, in 5 samples with PSCA mRNA overall scores of 3–6 and in 2 with scores of 9 there were less or negative PSCA protein expression (i.e. scores of 0–4), suggesting that this may reflect posttranscriptional modification of PSCA or that the epitopes recognized by PSCA mAb may be obscured in some cancers. The data demonstrate that the results of PSCA immunostaining were consistent with those of mRNA ISH analysis, showing a high degree of correlation between PSCA protein and mRNA expression.

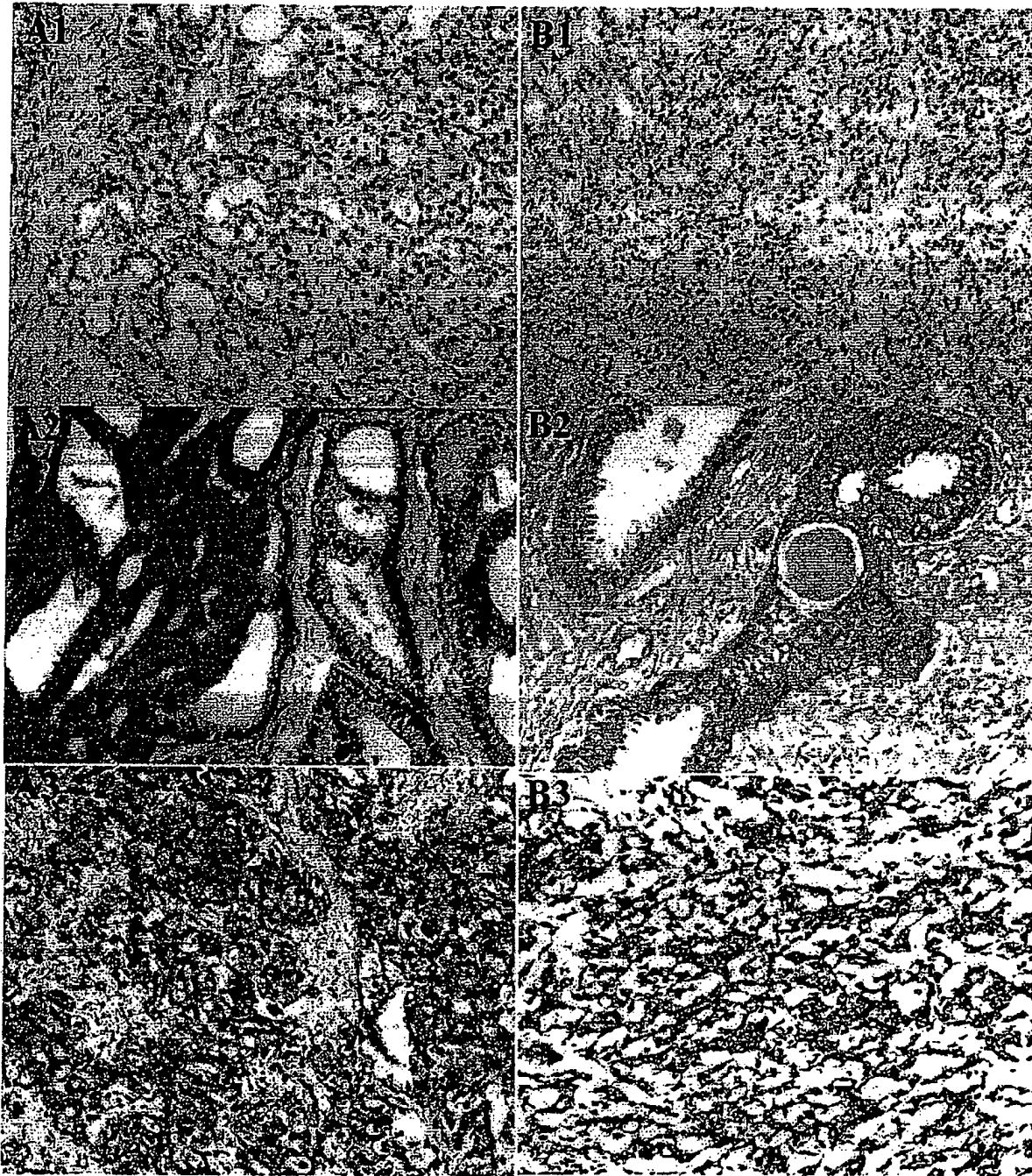


Figure 1

Representatives of PSCA IHC and ISH staining in Pca (A. IHC staining, B. ISH staining, $\times 200$ magnification). A₁, B₁: negative control of IHC and ISH. PBS replacing the primary antibody (A₁) and hybridization with a sense PSCA probe (B₁) showed no background staining. A₂, B₂: a moderately differentiated Pca (Gleason score = $3+3 = 6$) with moderate staining (composite score = 6) in all malignant cells; A₂: IHC shows not only cell surface but also apparent cytoplasmic staining of PSCA protein. A₃, B₃: a poorly differentiated Pca (Gleason score = $4+4 = 8$) with very strong staining (composite score = 9) in all malignant cells.

Discussion

PSCA is homologous to a group of cell surface proteins that mark the earliest phase of hematopoietic development. PSCA mRNA expression is prostate-specific in normal male tissues and is highly up-regulated in both androgen-dependent and-independent Pca xenografts (LAPC-4 tumors). We hypothesize that PSCA may play a role in Pca tumorigenesis and progression, and may serve as a target for Pca diagnosis and treatment. In this study, IHC and ISH showed that in general there were weak or absent PSCA protein and mRNA expression in BPH and low grade PIN tissues. However, PSCA protein and mRNA are widely expressed in HGPIN, the putative precursor of invasive Pca, suggesting that up-regulation of PSCA is an early event in prostate carcinogenesis. Recently, Reiter RE et al [1], using ISH analysis, reported that 97 of 118 (82%) HGPIN specimens stained strongly positive for PSCA mRNA. A very similar finding was seen on mouse PSCA (mPSCA) expression in mouse HGPIN tissues by Tran C. P et al [8]. These data suggest that PSCA may be a new marker associated with transformation of prostate cells and tumorigenesis.

We observed that PSCA protein and mRNA are highly expressed in a large percentage of human prostate cancers, including advanced, poorly differentiated, androgen-independent and metastatic cases. Fluorescence-activated cell sorting and confocal/ immunofluorescent studies demonstrated cell surface expression of PSCA protein in Pca cells [9]. Our IHC expression analysis of PSCA shows not only cell surface but also apparent cytoplasmic staining of PSCA protein in Pca specimens (Figure 1). One possible explanation for this is that anti-PSCA antibody can recognize PSCA peptide precursors that reside in the cytoplasm. Also, it is possible that the positive staining that appears in the cytoplasm is actually from the overlying cell membrane [5]. These data seem to indicate that PSCA is a novel cell surface marker for human Pca.

Our results show that elevated level of PSCA expression correlates with high grade (i.e. poor differentiation), increased tumor stage and progression to androgen-independence of Pca. These findings support the original IHC analyses by Gu Z et al [9], who reported that PSCA protein expressed in 94% of primary Pca and the intensity of PSCA protein expression increased with tumor grade, stage and progression to androgen-independence. Our results also collaborate the recent work of Han KR et al [10], in which the significant association between high PSCA expression and adverse prognostic features such as high Gleason score, seminal vesicle invasion and capsular involvement in Pca was found. It is suggested that PSCA overexpression may be an adverse predictor for recurrence, clinical progression or survival of Pca. Hara H et al [11] used RT-PCR detection of PSA, PSMA and PSCA in 1

ml of peripheral blood to evaluate Pca patients with poor prognosis. The results showed that among 58 Pca patients, each PCR indicated the prognostic value in the hierarchy of PSCA>PSA>PSMA RT-PCR, and extraprostatic cases with positive PSCA PCR indicated lower disease-progression-free survival than those with negative PSCA PCR, demonstrating that PSCA can be used as a prognostic factor. Dubey P et al [12] reported that elevated numbers of PSCA + cells correlate positively with the onset and development of prostate carcinoma over a long time span in the prostates of the TRAMP and PTEN +/- models compared with its normal prostates. Taken together with our present findings, in which PSCA is overexpressed from HGPIN to almost frank carcinoma, it is reasonable and possible to use increased PSCA expression level or increased numbers of PSCA-positive cells in the prostate samples as a prognostic marker to predict the potential onset of this cancer. These data raise the possibility that PSCA may have diagnostic utility or clinical prognostic value in human Pca.

The cause of PSCA overexpression in Pca is not known. One possible mechanism is that it may result from PSCA gene amplification. In humans, PSCA is located on chromosome 8q24.2 [1], which is often amplified in metastatic and recurrent Pca and considered to indicate a poor prognosis [13-15]. Interestingly, PSCA is in close proximity to the c-myc oncogene, which is amplified in >20% of recurrent and metastatic prostate cancers [16,17]. Reiter RE et al [18] reported that PSCA and MYC gene copy numbers were co-amplified in 25% of tumors (five out of twenty), demonstrating that PSCA overexpression is associated with PSCA and MYC coamplification in Pca. Gu Z et al [9] recently reported that in 102 specimens available to compare the results of PSCA immunostaining with their previous mRNA ISH analysis, 92 (90.2%) had identically positive areas of PSCA protein and mRNA expression. Taken together with our findings, in which we detected moderate to strong expression of PSCA protein and mRNA in 34 of 40 (85%) Pca specimens examined simultaneously by IHC and ISH analyses, it is demonstrated that PSCA protein and mRNA overexpressed in human Pca, and that the increased protein level of PSCA was resulted from the upregulated transcription of its mRNA.

At present, the regulation mechanisms of human PSCA expression and its biological function are yet to be elucidated. PSCA expression may be regulated by multiple factors [18]. Watabe T et al [19] reported that transcriptional control is a major component regulating PSCA expression levels. In addition, induction of PSCA expression may be regulated or mediated through cell-cell contact and protein kinase C (PKC) [20]. Homologues of PSCA have diverse activities, and have themselves been involved in

carcinogenesis. Signalling through SCA-2 has been demonstrated to prevent apoptosis in immature thymocytes [21]. Thy-1 is involved in T cell activation and transduces signals through src-like tyrosine kinases [22]. Ly-6 genes have been implicated both in tumorigenesis and in cell-cell adhesion [23-25]. Cell-cell or cell-matrix interaction is critical for local tumor growth and spread to distal sites. From its restricted expression in basal cells of normal prostate and its homology to SCA-2, PSCA may play a role in stem/progenitor cell function, such as self-renewal (i.e. anti-apoptosis) and/or proliferation [1]. Taken together with the results in the present study, we speculate that PSCA may play a role in tumorigenesis and clinical progression of Pca through affecting cell transformation and proliferation. From our results, it is also suggested that PSCA as a new cell surface antigen may have a number of potential uses in the diagnosis, therapy and clinical prognosis of human Pca. PSCA overexpression in prostate biopsies could be used to identify patients at high risk to develop recurrent or metastatic disease, and to discriminate cancers from normal glands in prostatectomy samples. Similarly, the detection of PSCA-overexpressing cells in bone marrow or peripheral blood may identify and predict metastatic progression better than current assays, which identify only PSA-positive or PSMA-positive prostate cells.

In summary, we have shown in this study that PSCA protein and mRNA are maintained in expression from HGPIN through all stages of Pca in a majority of cases, which may be associated with prostate carcinogenesis and correlate positively with high tumor grade (poor cell differentiation), advanced stage and androgen-independent progression. PSCA protein overexpression is due to the upregulation of its mRNA transcription. The results suggest that PSCA may be a promising molecular marker for the clinical prognosis of human Pca and a valuable target for diagnosis and therapy of this tumor.

Competing interests

None declared.

References

- Reiter RE, Gu Z, Watabe T, Thomas G, Szigeti K, David E, Wahl M, Nisitani S, Yamashiro J, Le Beau MM, Loda M, Witte ON: Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer. *Proc Natl Acad Sci USA* 1998, 95:1735-1740.
- Gleason DF: Histologic grading and clinical staging of prostatic carcinoma. In: *Urologic Pathology: The Prostate* Edited by: Tannebaum M. Philadelphia, Lea & Febiger; 1977:171-197.
- Brawer MK: Prostatic intraepithelial neoplasia: a premalignant lesion. *Hum Pathol* 1992, 23:242-248.
- Amin MB, Ro JY, Ayala AC: Prostatic intraepithelial neoplasia: relationship to adenocarcinoma of prostate. *Pathol Annu* 1994, 29:1-30.
- Amara N, Palapattu GS, Schrage M, Gu Z, Thomas GV, Dorey F, Said J, Reiter RE: Prostate stem cell antigen is overexpressed in human transitional cell carcinoma. *Cancer Res* 2001, 61:4660-4665.
- Hanas JS, Lerner MR, Lightfoot SA, Raczowski C, Kastens DJ, Brackett DJ, Postier RG: Expression of the cyclin-dependent kinase inhibitor p21 (WAF1/CIP1) and p53 tumor suppressor in dysplastic progression and adenocarcinoma in Barrett esophagus. *Cancer (Phila)* 1999, 86:756-763.
- Egevad L, Gramfors T, Karlberg L: Prognostic value of the Gleason score in prostate cancer. *BJU Int* 2002, 89:538-542.
- Tran CP, Lin C, Yamashiro J, Reiter RE: Prostate stem cell antigen is a marker of late intermediate prostate epithelial cells. *Mol Cancer Res* 2002, 1:113-121.
- Gu Z, Thomas G, Yamashiro J, Shintaku IP, Dorey F, Raitano A, Witte ON, Said JW, Loda M, Reiter RE: Prostate stem cell antigen (PSCA) expression increases with high Gleason score, advanced stage and bone metastasis in prostate cancer. *Oncogene* 2000, 19:1288-1296.
- Han KR, Sellgren DB, Liu X, Horvath S, Shintaku PI, Thomas GV, Said JW, Reiter RE: Prostate stem cell antigen expression is associated with gleason score, seminal vesicle invasion and capsular invasion in prostate cancer. *J Urol* 2004, 171:1117-1121.
- Hara H, Kasahara T, Kawasaki T, Bilim V, Obara K, Takahashi K, Tomita Y: Reverse Transcription-Polymerase Chain Reaction Detection of Prostate-specific Antigen, Prostate-specific Membrane Antigen, and Prostate Stem Cell Antigen in One Milliliter of Peripheral Blood. *Clin Cancer Res* 2002, 8:1794-1799.
- Dubey P, Wu H, Reiter RE, Witte ON: Alternative pathways to prostate carcinoma activate prostate stem cell antigen expression. *Cancer Res* 2001, 61:3256-3261.
- Visa korpi T, Kallioniemi AH, Syvanen AC, Hyttinen ER, Karhu R, Tammela T, Isola JJ, Kallioniemi OP: Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res* 1995, 55:342-347.
- Sato K, Qian J, Slezak JM, Lieber MM, Bostwick DG, Bergstrahl EJ, Jenkins RB: Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma. *J Natl Cancer Inst* 1999, 91:1574-1580.
- Van Den Berg C, Guan XY, Von Hoff D, Jenkins R, Bittner J, Griffin C, Kallioniemi O, Visakorpi T, McGill J, Herath J, Epstein J, Sarosdy M, Meltzer P, Trent J: DNA sequence amplification in human prostate cancer identified by chromosome microdissection: potential prognostic implications. *Clin Cancer Res* 1995, 1:11-18.
- Jenkins RB, Qian J, Lieber MM, Bostwick DG: Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization. *Cancer Res* 1997, 57:524-531.
- Nupponen NN, Kakkola L, Koivisto P, Visakorpi T: Genetic alterations in hormone-refractory recurrent prostate carcinomas. *Am J Pathol* 1998, 153:141-148.
- Reiter RE, Sato I, Thomas G, Qian J, Gu Z, Watabe T, Loda M, Jenkins RB: Coamplification of prostate stem cell antigen (PSCA) and MYC in locally advanced prostate cancer. *Genes Chromosomes Cancer* 2000, 27:95-103.
- Watabe T, Lin M, Donjacour AA, Cunha GR, Witte ON, Reiter RE: Growth, regeneration, and tumorigenesis of the prostate activates the PSCA promoter. *Proc Natl Acad Sci USA* 2002, 99:401-406.
- Bahrenberg G, Brauers A, Joost HG, Jakse G: PSCA expression is regulated by phorbol ester and cell adhesion in the bladder carcinoma cell line RT112. *Cancer Lett* 2001, 168:37-43.
- Noda S, Kosugi A, Sakoh S, Narumiya S, Hamaoka T: Protection from anti-TCR/CD3-induced apoptosis in immature thymocytes by a signal through thymic shared antigen-1/stem cell antigen-2. *J Exp Med* 1996, 183:2355-2360.
- Thomas PM, Samelson LE: The glycosphosphatidylinositol-anchored Thy-1 molecule interacts with the p60lyn protein tyrosine kinase in T cells. *J Biol Chem* 1992, 267:12317-12322.
- Bamezai A, Rock KL: Overexpressed Ly-6A.2 mediated cell-cell adhesion by binding a ligand expressed on lymphoid cells. *Proc Natl Acad Sci USA* 1995, 92:4294-4298.
- Katz BZ, Eshel R, Sagl-Assif O, Witz IP: An association between high Ly-6A/E expression on tumor cells and a highly malignant phenotype. *Int J Cancer* 1994, 59:684-691.
- Brakenhoff RH, Gerretsen M, Knippels EM, van Dijk M, van Essen H, Weghuis DO, Sinke RJ, Snow GB, van Dongen GA: The human E48 antigen, highly homologous to the murine Ly-6 antigen ThB, is a GPI-anchored molecule apparently involved in keratinocyte cell-cell adhesion. *J Cell Biol* 1995, 129:1677-1689.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	: Botstein, et al.
Appl. No.	: 10/032,996
Filed	: December 27, 2001
For	: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	: Fredman, J.
Group Art Unit	: 1637

DECLARATION OF VICTORIA SMITH, Ph.D., UNDER 37 CFR §1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, Dr. Victoria Smith, declare and state as follows:

1. I am a Senior Scientist in the Department of Molecular Biology of Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080.
2. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
3. I joined Genentech in 1996. For approximately three years, I directed a laboratory in the Department of Molecular Biology. During this time I was involved in target discovery for the Tumor Antigen Project, using DNA microarrays to discover genes differentially expressed in tumors compared to their expression in normal tissues. In connection with the above-identified patent application, I directed the generation and analysis of the microarray data attached as Exhibit B.
4. Exhibit B reports the results of the microarray analysis conducted on the gene encoding PRO539 (DNA47465) as part of the investigation of several newly discovered DNA sequences. The column "Unq Id" identify the gene as 340, which is DNA47465, while the column "DNA Id" identifies the particular lot of PCR product used. The microarray experiments were performed using well-established and accepted microarray techniques known in the art. (See, e.g., Nature Revs. Genetics, 5:229-237 (2004), attached as Exhibit C). The DNA samples used in the microarray studies were obtained from individual lung tumor tissue samples or individual normal lung tissue samples. The individual tumor and normal lung samples were each

Appl. No. : 10/032,996
Filed : December 27, 2001

compared to pooled samples of normal epithelial tissue. The level of expression in the lung tumor or normal lung tissue sample was compared to the normal pooled epithelial sample, and reported as a raw ratio. The average of the normal lung samples was then used to normalize the data to generate a ratio of expression of the PRO539 gene in lung tumor samples compared to the average expression in normal lung tissue. In the results reported in Exhibit B, a ratio of 2.0 or greater is a significant result, and indicates a significant increase in expression of the PRO539 gene in lung tumor tissue compared to the normal lung tissue controls.

5. The results of the microarray studies reported in Exhibit B indicate that the gene encoding PRO539 (DNA47465) is significantly overexpressed in eight of the twenty-six lung tumor samples tested compared to the normal lung tissue controls. That is the equivalent of nearly one in every three samples (31%). In addition, four out of five squamous cell lung carcinomas (80%) are significantly overexpressed (shown in bold). In contrast, only one of the seven individual normal lung tissue samples shows significant overexpression of the PRO539 gene (14%).

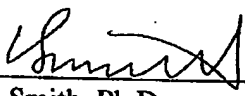
6. It is well-established in the art that overexpression of the mRNA for a gene is likely to lead to overexpression of the corresponding protein. Support for this statement can be found, for example, in the Molecular Biology of the Cell, a leading textbook in the field. (Bruce Alberts, *et al.*, Molecular Biology of the Cell (4th ed. 2002), excerpts submitted herewith as Exhibit D). Figure 6-3 on page 302 illustrates the basic principle that there is a correlation between increased gene expression and increased protein expression. The accompanying text states that "a cell can change (or regulate) the expression of each of its genes according to the needs of the moment – *most obviously by controlling the production of its mRNA.*" Molecular Biology of the Cell at 302, emphasis added. Similarly, figure 6-90 on page 364 illustrates the path from gene to protein. The accompanying text states that while potentially each step can be regulated by the cell, "the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes." Molecular Biology of the Cell at 364. This point is repeated on page 379, where the authors state that of all the possible points for regulating protein expression, "[f]or most genes transcriptional controls are paramount." Molecular Biology of the Cell at 379.

7. While not every lung tumor sample tested shows overexpression of the PRO539 gene, the data in Exhibit B indicate that a significant portion of lung tumors do (30% of all lung tumors tested and 80% of squamous cell tumors tested), while only one of the normal lung tissue samples shows overexpression (14%). Given the known correlation between overexpression of a gene and the corresponding overexpression of the encoded protein, it is very likely that a similar number of lung tumors will overexpress the PRO539 protein, while very few normal lung tissue samples likely will. Together with the data reported in Example 16 that the gene encoding PRO539 is amplified in some lung tumors, including squamous cell lung carcinoma, the results reported in Exhibit B indicate that the PRO539 gene and protein, as well as antibodies to the encoded protein, can be used to differentiate some cancerous lung tissue, particularly squamous cell carcinoma, from normal lung tissue. Because not all lung tumors show overexpression of PRO539, it cannot be used to exclude a sample being tested as non-cancerous. However, the PRO539 gene, protein, and corresponding antibodies are useful as a diagnostic tool for lung cancer, particularly squamous cell carcinoma, since a very high percentage of squamous cell lung

Appl. No. : 10/032,996
Filed : December 27, 2001

carcinomas overexpress the gene and most likely the encoded protein, while very few normal lung samples do.

8. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By:  Date: 01/20/05
Victoria Smith, Ph.D.

SADCSBSGBSG-1775.DOC
011905

EXHIBIT A

VICTORIA SMITH

Genentech Inc.
Dept. Molecular Biology
1 DNA Way
South San Francisco CA 94080
Ph: (650) 225 7382
Fax: (650) 225 6497
Email: victoria@gene.com

EDUCATION

Ph.D. (1991) Molecular Biology, University of Cambridge, Cambridge, United Kingdom.

Honors (1987) Biochemistry, University of Western Australia, Australia.

Bachelor of Science (1986) Physical and Inorganic Chemistry, and Biochemistry, University of Western Australia, Australia.

WORK AND RESEARCH EXPERIENCE

Senior Scientist, Genentech Inc (August 1996 - present; promoted to Senior Scientist March 2001)

Lab head, Dept. Molecular Biology

- Identification of potential therapeutic targets for cancer using novel microarray technology
- Discovery and identification of novel secreted proteins
- Development of cancer therapeutics

Stanford University, California, U.S.A. (February 1995 - August 1996)

Research Fellow, Department of Biochemistry

Research: Genomic functional analysis of chromosome V of *Saccharomyces cerevisiae*

Stanford University, California, U.S.A. (January 1992 - January 1995)

Postdoctoral Fellow, Department of Genetics.

Research: Development of new methodology for whole genome functional analysis in microorganisms using genomic sequence data and insertional mutagenesis.

Awards

Human Frontiers Science Program Organization Long Term Fellowship (accepted, 4/01/93 - 1/31/95).

American Cancer Society (California Division) Fellowship (1993, declined).

Cambridge University, United Kingdom (October 1988 - December 1991)

Research undertaken at the Medical Research Council Laboratory of Molecular Biology, Cambridge, UK, for the degree of Doctor of Philosophy, Cambridge University.

Thesis: A Molecular Genetic Analysis of Yeast Chromosome IX.

Thesis Advisor: Dr. Barclay Barrell.

Awards

Max Perutz Prize in 1991 for outstanding performance by a graduate student. Awarded for advances in genomic-scale DNA sequencing methodology, and genetic analysis of the *SNP1* gene of *Saccharomyces cerevisiae*.

King Edward Memorial Hospital for Women, Western Australia (1988)

Research Scientist.

Research: Analysis of hormone-inducible mRNAs in breast tumors.

University of Western Australia (1984 - 1987)

1984 - 1986: Bachelor of Science degree with double major in Biochemistry and Physical and Inorganic Chemistry.

1987: First Class Honors in Biochemistry. Thesis: Nuclease Sensitivity and Methylation Patterns of the Phenylalanine Hydroxylase Gene.

Awards

Association of Commonwealth Universities Scholarship for study in the United Kingdom (accepted, October 1988 - October 1991, University of Cambridge).

Hackett Scholarship for overseas study (1988, declined).

Lady James Prize in Natural Science in 1986: Best student completing Bachelor of Science degree with a major in a natural science.

JWH Lugg Prize in Biochemistry in 1986: Best student completing major in Biochemistry.

Convocation Prize in Science in 1985: Best student completing major in second year Physics, Geology or Chemistry.

Shell Prize in Chemistry in 1985: Best student completing major in second year Chemistry.

PUBLICATIONS

V. Smith, E.F. Shen, D. Wieand, T.H. Landon, N.A. Wong, A.M. Lessells, S. Paterson-Brown, T.D. Wu, J.Z. Tang, K.J. Hillan, I.D. Penman, Expression Analysis of the Metaplasia-Dysplasia-Carcinoma Sequence in Barrett's Esophagus (submitted).

Tice DA. Szeto W. Soloviev I. Rubinfeld B. Fong SE. Dugger DL. Winer J. Williams PM. Wieand D. Smith V. Schwall RH. Pennica D. Polakis P. *Journal of Biological Chemistry*. 277(16):14329-35, 2002 Apr 19.

N.J. Maughan, F. Lewis, V. Smith (2001) *Journal of Pathology* 195, 3-6.

F. Lewis, N.J. Maughan, V. Smith, K. Hillan, P. Quirke (2001) *Journal of Pathology* 195, 66-71.

D.J. Garfinkel, M.J. Curcio and V. Smith (1998) Ty Mutagenesis *Methods in Microbiology*, volume 26, 101-117.

C. Churcher, S. Bowman, K. Badcock, A. Bankier, D. Brown, T. Chillingworth, R. Connor, K. Devlin, S. Gentles, N. Hamlin, D. Harris, T. Horsnell, S. Hunt, K. Jagels, M. Jones, G. Lye, S. Moule, C. Odell, D. Pearson, M. Rajandream, P. Rice, N. Rowley, J. Skelton, V. Smith, S. Walsh, S. Whitehead & B. Barrell. (1997) *Nature*, 387, 84-87.

F. S. Dietrich, J. Mulligan, K. Hennessy, M. A. Yelton, E. Allen, R. Araujo, E. Aviles, A. Berno, T. Brennan, J. Carpenter, E. Chen, J. M. Cherry, E. Chung, M. Duncan, E. Guzman, G. Hartzell, S. Hunicke-Smith, R. W. Hyman, A. Kayser, C. Komp, D. Lashkari, H. Lew, D. Lin, D. Mosedale, K. Nakahara, A. Namath, R. Norgren, P. Oefner, C. Oh, F.X. Petel, D. Roberts, P. Sehl, S. Schramm, T. Shogren, V. Smith, P. Taylor, Y. Wei, D. Botstein & R. W. Davis. (1997) *Nature* 387, 78-81.

V. Smith, K. Chou, D. Lashkari, D. Botstein, and P. O. Brown. (1996). Functional Analysis of the Genes of Yeast Chromosome V by Genetic Footprinting. *Science* **274**, 2069-74

V. Smith, D. Botstein, and P. O. Brown (1995) Genetic Footprinting: A genomic strategy for determining a gene's function given its sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 6479-6483.

V. Smith, M. Craxton, A. T. Bankier, C. M. Brown, W. D. Rawlinson, M. Chee, and B. G. Barrell (1995) Microtiter methods for the preparation and fluorescent sequencing of M13 clones. *Recombinant DNA Methodology II: a volume in the Selected Methods in Enzymology Series*, pp. 607 - 621.

V. Smith, M. Craxton, A. T. Bankier, C. M. Brown, W. D. Rawlinson, M. S. Chee, and B. G. Barrell (1993) Microtiter methods for the preparation and fluorescent sequencing of M13 clones. *Methods in Enzymology*, **218**, 173-187.

V. Smith and M. S. Chee (1991) A simple method for sequencing the complementary strand of ssDNA from M13 clones. *Nucleic Acids Research* **19**, 6957.

V. Smith and B. G. Barrell (1991) Cloning of a Yeast U1 snRNP 70K protein homologue: functional conservation of an RNA binding domain between humans and yeast. *EMBO Journal* **10**, 2627-2643.

W. D. Rawlinson, M. S. Chee, V. Smith and B. G. Barrell (1991) Preparation of large numbers of single stranded DNA templates by rescue from phagemids in microtiter plates. *Nucleic Acids Research* **19**, 4779.

V. Smith, C. M. Brown, A. T. Bankier, and B. G. Barrell (1990) Semi-automated preparation of DNA templates for large scale sequencing projects. *DNA Sequence* **1**, 73-78.

S. J. Wysocki, E. Hahnel, S. P. Wilkinson, V. Smith, and R. Hahnel (1990) Hormone-sensitive gene expression in breast tumors. *Anticancer Research* **10**, 185-188

S. J. Wysocki, E. Hahnel, A. Masters, V. Smith, A. J. McCartney and R. Hahnel (1990)
Detection of pS2 messenger RNA in gynecological cancers. *Cancer Research* 50, 1800-1802

PATENTS

Genetic Footprinting: Insertional Mutagenesis and Genetic Selection. U.S. Patent No. 5,612,180. Inventors: Patrick Brown and Victoria Smith

PATENTS FILED (at Genentech Inc.)

Methods of Detecting and Quantifying Gene Expression. Inventors: Victoria Smith, Edward Robbie, David Lowe, James Marsters.

Compositions and Methods for the Treatment of Cancer. Inventors: Victoria Smith, Austin Gurney, Audrey Goddard, Fred DeSavauge

Diagnostic for Dysplasia in Barrett's Esophagus. Inventor: Victoria Smith

Numerous composition of matter filings related to novel gene discovery, pending

EXHIBIT B

Unq Id	DNA Id	Experiment Name	Raw Ratio (sample/pooled epithelial)	Normalized Ratio (sample/normal lung)
--------	--------	-----------------	--	--

Lung Tumor Samples

340.	208,278.	lung tumor R1057 vs epi pool	3.429	2.463
340.	208,278.	lung tumor R1094 vs epi pool	1.716	1.232
340.	208,278.	lung tumor R1372 vs epi pool	0.046	0.033
340.	208,278.	lung tumor R1417 vs epi pool	1.106	0.794
340.	208,278.	lung tumor R1431 vs epi pool	0.242	0.174
340.	208,278.	lung tumor R419 vs epi pool	1.413	1.015
340.	208,278.	lung tumor R542 vs epi pool	0.429	0.308
340.	208,278.	lung tumor R544 vs epi pool	0.735	0.528
340.	208,278.	lung tumor R685 vs epi pool	0.247	0.177
340.	208,278.	lung tumor R688 vs epi pool	0.897	0.644
340.	208,278.	lung tumor R693 vs epi pool	0.755	0.542
340.	208,278.	lung tumor R737 vs epi pool	1.508	1.083
340.	208,278.	lung tumor R742 vs epi pool	0.313	0.225
340.	295,413.	lung tumor	3.213	2.307
340.	208,278.	lung tumor R543 vs epi pool	0.402	0.289
340.	208,278.	lung tumor1604/REF.RNA	1.06	0.761
340.	208,278.	lung tumor1683	11.523	8.275
340.	208,278.	lung tumor1853	0.067	0.048
340.	208,278.	lung tumor2454/ref.RNA	11.028	7.920
340.	208,278.	LungCaSq-hf1602	1.103	0.792
340.	208,278.	LungCaSq-hf1647	2.794	2.007
340.	208,278.	LungSqCa-hf1293	4.728	3.396
340.	208,278.	LungSqCa-hf1646	5.399	3.877
340.	208,278.	LungSqCa-hf?	6.306	4.529
340.	208,278.	100ngLungBaCa1069 vs 25ngEpi1409	0.376	0.270
340.	208,278.	100ngLungBaCa994 vs 25ngEpi1409	1.706	1.225

Normal Lung Samples

340.	208,278.	N. lung R1415 vs univ.ref	0.433	0.311
340.	208,278.	N. lung R1431 vs ref.	0.567	0.407
340.	208,278.	N. lung R417 vs univ. ref	1.372	0.985
340.	208,278.	1054LungInflamTA2	1.585	1.138
340.	208,278.	1415LungInflamTA2	1.408	1.011
340.	208,278.	423LungInflamTA2	3.709	2.664
340.	208,278.	488-552InflmdLungTA2	0.673	0.483

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.